



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Investigation of alder (Alnus incana)
chloroplast ribosomal RNA genes

By

Madeleine Lévesque

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for
the M.Sc. degree in Biology

University of Ottawa
Ottawa, Ontario, Canada

© Madeleine Lévesque, Ottawa, Canada, 1990



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-60100-0

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGEMENTS

I thank Dr. D.A. Johnson for his guidance, encouragement, advice during the course of this project. In addition, I wish to thank Dr. M. McBurney and Dr. B. Miki for having served on my research committee and Dr. Linda Bonen for helpful advice.

I acknowledge financial support from N.S.E.R.C., the Biology Department at the University of Ottawa and Dr. D.A. Johnson.

I would like to thank the "bodies" in the lab for being there and for making the lab a pleasant place. I also thank Mr. P. Brunon and Mr. J. Hélie for the speedy processing of the figures presented in this work. Thanks are also extended to Dr. J. Hattori at Agriculture Canada for his help with the sequence analysis.

I also want to thank the people at CISTI especially S.L., A.R., P.L. for all their help to locate the bulk of the present material. Merci beaucoup.

Most importantly, I wish to thank my parents, my sister, Michel and his parents, for the love, support and encouragement they have provided throughout this project. It would have been impossible to make it without you, and I love you dearly. D.M.

ABSTRACT

Chloroplasts are intracellular organelles responsible for plant photosynthesis. These organelles contain their own genome; chloroplast DNA is found as a circular double-stranded molecule, ranging in size between 120-160 kilobase pairs (Sugiura, 1989a). Within this genome are encoded genes for the genetic apparatus eg. transfer RNA genes (*trn*), ribosomal protein genes (*rpl* or *rps*) and ribosomal RNA genes (16S, 23S, 4.5S, 5S). Other genes include those for the photosynthetic apparatus eg. the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/ oxygenase (RuBisCo) (*rbcL*), photosystem I genes (*psa*), photosystem II genes (*psb*), as well as cytochrome b/f complex genes, ATP synthase genes. In addition many open reading frames have been identified by sequence analysis, some may be putative genes. Thus many of the genes required for the maintenance and function of this organelle are found encoded on its genome. The remaining gene products are encoded by nuclear genes.

Translation of the chloroplast genes, as is the case for any other gene, is the key step in gene expression at which the genetic information encoded in the genes is converted into proteins which ultimately determine an organism's phenotype (Raué *et al.*, 1988; Noller *et al.*, 1987). The common function of all ribosomes is clearly reflected in a common basic anatomy (Raué *et al.*, 1988);

therefore, ribosomal RNAs which are components within these ribonucleoprotein particles are molecules of choice for relatedness studies. Ribosomal RNA genes have long been useful to determine phylogenetic relationships between organisms. Some 400 genes have been sequenced for the small subunit rRNAs and close to 70 for the large subunit rRNA (Gutell and Woese, 1990) which represent organisms forming the three major lines of descent: eubacterial, eukaryotic and archaebacterial. No ribosomal RNA gene has yet been isolated from a tree species. In an effort to position the alder, Alnus incana, in the scope of phylogenetic relationships, ribosomal RNA genes from its chloroplasts have been cloned and analysed.

A genomic library was constructed in the vector LambdaGEM-11 with DNA isolated from leaf tissue. Plaques were screened using radiolabeled probes to isolate positive clones. DNA was isolated from these clones and subjected to restriction analysis to identify fragments containing 23S and 16S rRNA gene sequences. These fragments were then subcloned into an M13 vector and overlapping deletion clones were generated by the Dale procedure (Dale et al., 1985); they were also subcloned into a plasmid vector to allow further mapping and sequencing. DNA fragments covering the length of the 23S and 16S rRNA genes as well as trnI and sections of trnA were sequenced.

The order of chloroplast genes within the sequenced region of alder is identical to that seen in higher plants: 16S-trnI-trnA-23S (Palmer *et al.*, 1988a). Comparisons of the primary sequence of the alder 16S and 23S rRNA genes with available ribosomal RNA and DNA sequences have demonstrated the greatest degree of homology between the alder sequences and those of other higher plant sequences, such as those of maize and tobacco sequences and a lesser degree, to those of lower plants such as liverwort. Comparisons to bacterial sequences yield higher homologies to the cyanobacterial sequences as opposed to those of *E. coli*, which was anticipated from the current thought on the origin of the chloroplast within eucaryotic cells. Further sequence analysis shows that the primary sequence obtained for the 16S rRNA gene in alder can be folded to follow the secondary structure of the maize 16S rRNA molecule. This further illustrates the conserved nature of this molecule's secondary structure.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
TABLE OF CONTENTS	vii
LIST OF FIGURES	x
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER 1: INTRODUCTION	1
I. Chloroplasts	1
A. Chloroplast development	2
B. Chloroplast genome	4
1. Genome structure	4
2. Genes encoded on the chloroplast genome	5
C. Ribosomal RNA genes	13
D. Transfer RNA genes	18
II. Use of ribosomal RNA as models for evolutionary studies	20
A. Primary sequence of rRNAs	21
B. Secondary structure of rRNAs	22
III. Alder as a model system for symbiosis	30
IV. Objectives of this work	33
CHAPTER 2: MATERIALS AND METHODS	36
I. Isolation of DNA and RNA	36
A. DNA isolation	36
1. Isolation of genomic DNA from plants....	36
2. Small scale DNA preparation from positive lambda phages	36
a) Preparation of the lysate	36
b) Extraction of DNA	37
3. Large scale DNA preparation from lambda phage	38
a) Preparation of the lysate	38
b) Extraction of DNA	38
4. Isolation of restriction fragments from agarose gels	39
5. Isolation of plasmid and M13 phage DNAs on a small scale	39
a) Plasmid DNA	39
b) Phage DNA	39
i) DIGE	39
ii) Preparation of the DNA .	40

3.	RNA isolation	40
	Alder RNA isolation	40
II.	Southern and Northern transfers, plaque lifts	41
A.	Southern transfers	41
1.	Standard procedure	41
2.	Genomic transfers	41
3.	Dot blot controls	42
B.	Northern transfers	42
III.	Probe preparation	43
IV.	Hybridizations	43
A.	Hybridizations and washes	43
1.	Southern transfers	43
2.	Northern transfers	44
B.	Rehybridizations	44
V.	Alder genomic library construction	45
A.	Isolation of MboI fragments for cloning	45
B.	Ligation of MboI fragments to LambdaGEM-11 vector	45
VI.	Library screening	46
A.	Plating of library	46
B.	Screening of genomic library	47
C.	Plaque purification	47
D.	Storage of the library	48
VII.	Subcloning into pTZ18R and M13mp19 vectors	48
A.	Vector preparations	48
B.	Ligations and transformations	48
VIII.	DNA sequencing	49
A.	Preparation of templates	49
1.	Template orientation	49
2.	Dale deletions	49
B.	Sequencing reactions	50
IX.	Data Analysis	51
X.	Growth and maintenance of alders	51
A.	Growing <i>Frankia</i> ACN 14a	51
B.	Sprouting alders (<i>Alnus incana</i>)	52
C.	Planting and inoculation of alders	52
D.	Collecting tissues	53
XI.	Bacterial strains and relevant phenotypes	54
X.	Buffers and other solutions	54
CHAPTER 3: RESULTS		57
I.	Source of probes	57
II.	Isolation of alder and birch DNAs	57
III.	Southern transfer analyses	59
	Alder (<i>Alnus incana</i>) Southern transfers	59
IV.	Alder genomic library construction	62
A.	Choice of restriction enzyme for partial digestion of genomic DNA	62
B.	Isolation of MboI fragments for cloning	63
C.	Library construction	64
1.	Choice of vector	64

2.	Characterization of the library	64
D.	Choice of host strain	66
E.	Library screening	67
1.	Initial screening	67
2.	Preliminary restriction analysis of isolated clones	68
V.	Restriction analysis of clones pursued in this work	71
VI.	Subcloning and sequencing	79
A.	Subcloning into pTZ18R and M13mp19	79
1.	Strategy of cloning	79
2.	Choice of vectors	79
3.	Nomenclature of subclones	80
4.	Confirmation of clone identity	81
B.	Sequencing of cloned fragments	81
VII.	DNA sequence analysis	89
A.	Nucleotide sequence evaluation	92
B.	Primary nucleotide sequence analysis: comparisons to other organisms	115
1.	16S rDNA	116
2.	trnI, trnA	119
3.	23S rDNA	123
C.	Conserved secondary structure	125
CHAPTER 4: DISCUSSION		131
I.	Alder sequences and phylogeny	131
II.	Use of cloned sequences as probes to differentiate between cellular DNA from different compartments	138
III.	Future work	145
A.	Short term	145
B.	Long term	146
REFERENCES		149
APPENDIX 1		165
APPENDIX 2		169
APPENDIX 3		178

LIST OF FIGURES

- Figure 1 p. 7 Complete map of the tobacco chloroplast genome
- Figure 2 p. 24 Phylogenetic trees prepared from small subunit and large subunit ribosomal RNA molecules
- Figure 3 p. 27 Secondary structural diagrams for the E. coli 16S and 23S rRNAs
- Figure 4 p. 60 Hybridization of alder DNA to cDNA insert probes
- Figure 5 p. 69 Alder genomic library screening
- Figure 6 p. 75 Restriction map of the alder clones λ Anod4-1 and λ Anod4-2 and their positions with respect to the tobacco chloroplast map
- Figure 7 p. 77 Linkage of two alder genes
- Figure 8 p. 83 Cloning strategy and subclone identification
- Figure 9 p. 85 Sequencing strategy for subclones derived from λ Anod4-1
- Figure 10 p. 94 Nucleotide sequences of CP516, CPTRN AND CP423 and sequence comparison of alder to maize and tobacco chloroplast 16S rRNA gene sequences
- Figure 11 p. 127 Secondary structure of the alder 16S rRNA
- Figure 12 p. 134 Comparison of the secondary structure of the 16S rRNA molecule among bacteria, mitochondria and chloroplast
- Figure 13 p. 172 Hybridization pattern of λ Anod5-5 to pAnod5 insert probe
- Figure 14 p. 174 Restriction map of the λ Anod5-5 clone

LIST OF FIGURES (CONTINUED)

- Figure 15 p. 176 Nucleotide sequence of subclone MBg50 and sequence comparison to maize mitochondrial 18S rRNA gene sequence
- Figure 16 p. 181 Hybridization of birch DNA to cDNA insert probes

LIST OF TABLES

- Table 1 p. 9 Genes encoded on the chloroplast genome
- Table 2 p. 65 Preparation of the alder genomic libraries: cloning summary
- Table 3 p. 113 G+C content of the nucleotide sequence obtained for the alder clone λ Anod4-1
- Table 4 p. 118 Comparison of the rrn promoter consensus sequences
- Table 5 p. 183 Hybridizing fragment sizes observed for alder and birch DNA to the three cDNA insert probes

LIST OF ABBREVIATIONS USED IN THIS WORK

A:	adenine
bp:	base pair(s)
cDNA:	complimentary DNA; DNA copy
C:	cytosine
dATP:	deoxyadenosine triphosphate
dCTP:	deoxycytosine triphosphate
DNA:	deoxyribonucleic acid
EDTA:	ethylene diamine tetraacetic acid
g:	gram
G:	guanine
kb:	kilobase pair(s)
l:	litre
m:	metre
mRNA:	messenger RNA
n:	nucleotide(s)
O.D.:	optical density
°C:	degree Celcius
pfu:	plaque forming units
RNA:	ribonucleic acid
rRNA:	ribosomal RNA
SDS:	sodium dodecylsulphate
T:	thymine
tRNA:	transfer RNA
UV:	ultraviolet light

CHAPTER 1: INTRODUCTION

I. CHLOROPLASTS

Chloroplasts are the photosynthetic organelles in plant and algal cells (Ohyama et al., 1988) which play a central role in photosynthesis. These structures contain the entire enzymatic machinery for this process. It is through photosynthesis that solar energy in the form of light is captured and converted to chemical energy. The study of non-Mendelian inheritance of mutants in chloroplast characters, at the turn of the century, suggested the existence of a separate genetic system in chloroplasts (Ohyama et al., 1988). It has been demonstrated that organelles have their own genome which can be isolated and used for further study.

Chloroplasts, as the organelle site of photosynthetic CO₂ fixation, ATP synthesis, and O₂ evolution are unique to algae and plants. Based on their central importance in photosynthesis as well as their developmental and differentiation capacities during plant development, plastids have become an important model system for the study of the regulation of photosynthetic gene expression and the interaction of nuclear and organelle genomes (Gruissem et al., 1989c).

The study of chloroplasts has focused on a number of areas. Targeting of proteins from cytosol to this organelle has received much attention due to the metabolic importance

of chloroplasts (location of fatty acid synthesis, starch metabolism, essential aspects of amino acid synthesis, as well as other processes) (Keegstra, 1989) as well as the complexity of their structure. Chloroplasts are separated into different compartments: outer membrane envelope, inner membrane space, inner membrane envelope, stroma, thylakoid membrane and thylakoid lumen. In addition, as the chloroplast genome has a limited protein coding capacity (Weisbeek *et al.*, 1987), nuclear-encoded proteins synthesized in the cytosol (eg. small subunit of RuBisCo (Dean *et al.*, 1989)), must be transported to and translocated across the chloroplast membranes (Keegstra and Olsen, 1989) and assembled into multimeric complexes (Hemmingsen *et al.*, 1988). Chloroplasts also are developmentally complex as described later. The genes encoded by the chloroplast genomes show novel modes of regulation with respect to development and light (Deng *et al.*, 1989; Gruissem *et al.*, 1989c; Kuhlemeier *et al.*, 1989; Link, 1988), and a trans-splicing mechanism for intron splicing had been elucidated for the chloroplast ribosomal protein gene rps12 (Gruissem, 1989b; Sugiura *et al.*, 1987).

A. Chloroplast development

In developing plants, chloroplasts are derived from undifferentiated precursor plastids (small spherical progenitor organelles) termed proplastids, or from

etioplasts in seedlings which are grown in the dark (Kirk and Tilney-Basset, 1978). Proplastids normally originate maternally during the formation of plant zygotes and are maintained in an undifferentiated state in meristematic cells of developing plants (Mullet, 1988; Gruissem, 1989). Proplastids may be formed during seasonal dormancy or maturation of specialized organs (Mullet, 1988). The prominent inner membrane, the thylakoid, found in mature chloroplasts is nearly absent in proplastids: these organelles contain low amounts of plastid DNA, RNA, ribosomes and soluble proteins (Mullet, 1988).

Plastid size and composition also change during chloroplast development. During the development of chloroplasts in photosynthetic tissues, photosynthetic electron-transfer components are assembled into Photosystem I and II, cytochrome *b₆f* and ATP synthase complexes each of which consists of up to 20 polypeptides (Gruissem, 1989c; Piechulla *et al.*, 1986). Proplastids and chloroplasts can also give rise to specialized plastid types which acquire other functions in non-photosynthetic plant organs. These structures include: etioplasts (chloroplast precursors in dark grown plants), and plastids specialized in the synthesis and accumulation of carotenoids (chromoplasts in fruit and flower), starch (amyloplasts, in root and tuber), terpenoids or lipids (Mullet, 1988).

It was found that the organization and composition of

DNA within photosynthetic and non-photosynthetic plastids are similar. This supports the idea of interconversion of plastid populations (Gruissem et al., 1989c; Mullet, 1988; Aguetaz et al., 1987). It is proposed from studies of induced specific biosynthetic pathways found in specialized nongreen plastids that biosynthetic pathways required for chloroplast biogenesis can be induced independently of chloroplast development (Mullet, 1988; Gruissem et al., 1989c).

B. Chloroplast genome

1. Genome structure

Chloroplasts contain their own autonomously replicating genome (Ohyama, et al., 1986b). In vascular plants, it is found as a double-stranded circular DNA molecule which varies in size between 120-160 kb (Sugiura, 1989a), compared to plant mitochondrial genome sizes which range between 200-2000 kb (Newton, 1988). The DNA molecules are associated with the inner membrane of the envelope or with internal thylakoid membranes in aggregates of 10-20 DNA molecules (Hermann and Possingham, 1980) or found in rings or in a more dispersed arrangement within the plastid (Miyamura et al., 1986). In wheat, the number of DNA copies per plastid increase during chloroplast development (Baumgartner et al., 1989; Miyamura et al., 1986).

The interesting feature of chloroplast DNAs in most

plants is that they contain two large inverted repeat (IR) regions (Sugiura, 1989a) separated by a large single copy region (LSC) and a small single copy (SSC) region (Figure 1). To date, only the chloroplast DNAs from pea (Pisum sativum) and broad bean (Vicia faba) have no inverted repeat; mung bean (Vigna radiata), common bean (Phaseolus vulgaris) and soybean (Glycine max) have retained one ancestral large inverted repeat (Palmer, et al., 1988b). Those legume chloroplast DNAs that have an ancestral large inverted repeat also share a highly conserved sequence order that differs by only a single inversion from the arrangement typical of most vascular plants (Palmer and Stein, 1986). Euglena gracilis, in contrast, contains three tandem repeats each having an rRNA gene cluster (Gray and Hallick, 1978).

2. Genes encoded on the chloroplast genome

Chloroplast DNAs contain all the chloroplast rRNA genes (4 genes in higher plants), tRNA genes (~30 genes) and all the genes for proteins synthesized in the chloroplast (100-150 genes) (Sugiura, 1989; Gray et al., 1984). Chloroplast genes from a number of higher plants such as maize (Dormann-Przybyl et al., 1986; Strittmatter et al., 1985; Schwarz et al., 1981), spinach (Tahar et al., 1986), pea (Palmer et al., 1988b; Jorgensen et al., 1987), soybean (Grabau, 1985; Morgens et al., 1984), tobacco (Tohdoh et al., 1981; Shinozaki et al., 1986), rice (Hiratsuka et al., 1989; Moon

et al., 1988), of lower plants such as liverwort (Ohyama et al., 1986; Fukuzawa et al., 1988; Kohchi et al., 1988; Ohyama et al., 1988) and unicellular organisms like Euglena (Mc Garvey and Helling, 1989), Anacystis nidulans (Kumano et al., 1986; Tomioka and Sugiura, 1983; Tomioka et al., 1981) and Chlamydomonas (Blowers et al., 1989; Lemieux et al., 1989; Rochaix et al., 1985; Rochaix, 1981) have been extensively studied with respect to genome structure and expression.

The complete nucleotide sequences for three chloroplast genomes have been determined: that of tobacco (Nicotiana tabacum: 155,844 bp, Shinozaki, et al., 1986a,b), liverwort (Marchantia polymorpha: 121,024 bp, Ohyama, et al., 1986) and rice (Oryza sativa: 134,525 bp, Hiratsuka, et al., 1989) and genes have been assigned to these sequences by their homologies to other already published sequences.

Genes encoded by the chloroplast genome generally are of two types: genes for the genetic apparatus and genes for the photosynthetic apparatus. In addition many open reading frames have been identified by sequence analysis; some may be putative genes. These are summarized in Table 1, modified from Sugiura, 1989a. Both strands of the chloroplast genome are actively expressed; it was documented that transcription switches strands over 30 times (Palmer et al., 1988a).

Photosynthesis, together with other plastid functions, requires the products of several hundred genes of which only

FIGURE 1: COMPLETE MAP OF THE TOBACCO CHLOROPLAST GENOME

The tobacco genome is a 155 844 bp molecule containing three regions: the LSC, long single copy; the SSC, the short single copy; the IR inverted repeat regions (dark boxes). The DNA strand which codes for the large subunit of ribulose-1,5-bisphosphate carboxylase has been designated as A and the complementary strand as B. The genes listed on the inside of the molecule are coded by strand B and are transcribed clockwise; those listed on the outside of the molecule are coded by the A strand and are transcribed counterclockwise. Asterisks indicate split genes. The figure was taken from Sugiura et al. 1987.

TABLE 1: GENES ENCODED ON THE CHLOROPLAST GENOME

The table lists the genes that have been identified on the chloroplast genome of most higher organisms. These have been separated in two groups: those necessary for the genetic apparatus and those required for the photosynthetic apparatus. This table was adapted from Sugiura 1989a, with a few modifications.

Genes for the genetic apparatus

rRNA genes

23S rDNA 23S rRNA
16S rDNA 16S rRNA
7S rDNA 7S rRNA
5S rDNA 5S rRNA
4.5SrDNA 4.5SrRNA
3S rDNA 3S rRNA

tRNA genes

trnA-UGC Ala-tRNA(UGC)
trnR-ACG Arg-tRNA(ACG)
trnR-UCU Arg-tRNA(UCU)
trnR-CCG Δ Arg-tRNA(CCG)
trnN-GUU Asn-tRNA(GUU)
trnD-GUG Asp-tRNA(GUC)
trnC-GCA Cys-tRNA(GCA)
trnQ-UUG Gln-tRNA(UUG)
trnE-UUC Glu-tRNA(UUC)
trnG-GCC Gly-tRNA(GCC)
trnG-UCC Gly-tRNA(UCC)
trnH-GUG His-tRNA(GUG)
trnI-GAU Ile-tRNA(GAU)

50S ribosomal proteins

rpl12 protein CL2
rpl14 protein CL14
rpl16 protein CL16
rpl20 protein CL20
rpl21 Δ protein CL21
rpl22 protein CL22
rpl23 protein CL23
rpl33 protein CL33
rpl36 protein CL36

RNA polymerase, subunits

rpoA subunit α
rpoB subunit β
rpoC subunit β

tufA Elongation factor Tu
infA Initiation factor 1

Genes for the photosynthesis apparatus

rbcL RuBisCo, large subunit

psaA PSI, P700 apoprotein A1
psaB PSI, P700 apoprotein A2

trnI-GAU	Ile-tRNA(CAU)	psaC	PSI, 9-kDa polypeptide
trnL-UAA	Leu-tRNA(UAA)		
trnL-CAA	Leu-tRNA(CAA)	psbA	PSII, D1 protein
trnL-UAG	Leu-tRNA(UAG)	psbB	PSII, 47kDa polypeptide
trnK-UUU	Lys-tRNA(UUU)	psbC	PSII, 43kDa polypeptide
trnM-CAU	fMet-tRNA(CAU)	psbD	PSII, D2 protein
trnM-CAU	Met-tRNA(CAU)	psbE	PSII, cytochrome b ₅₅₉ , (8 kDa)
trnF-GAA	Phe-tRNA(GAA)		
trnP-UGG	Pro-tRNA(UGG)	psbF	PSII, cytochrome b ₅₅₉ , (4 kDa)
trnS-GJA	Ser-tRNA(GGA)		
trnS-UGA	Ser-tRNA(UGA)	psbG	PSII, G protein
trnS-GCU	Ser-tRNA(GCU)	psbH	PSII, 10-kDa phosphoprotein
trnT-GGU	Thr-tRNA(GGU)		
trnT-UGU	Thr-tRNA(UGU)		
trnW-CCA	Trp-tRNA(CCA)	petA	b/f complex, cytochrome f
trnY-GUA	Tyr-tRNA(GUA)		
trnV-GAC	Val-tRNA(GAC)	petB	b/f complex, cytochrome b ₆
trnV-UAC	Val-tRNA(UAC)		
		petD	b/f complex, subunit IV

30S ribosomal proteins

		atpA	H ⁺ -ATPase, subunit α
rps2	protein CS2	atpB	H ⁺ -ATPase, subunit β
rps3	protein CS3	atpE	H ⁺ -ATPase, subunit ε
rps4	protein CS4	atpF	H ⁺ -ATPase, subunit I
rps7	protein CS7	atpH	H ⁺ -ATPase, subunit III
rps8	protein CS8	atpI	H ⁺ -ATPase, subunit IV
rps11	protein CS11		

rps12	protein CS12	ndhA	NADH dehydrogenase, ND1
rps14	protein CS14	ndhB	NADH dehydrogenase, ND2
rps15	protein CS15	ndhC	NADH dehydrogenase, ND3
rps16	protein CS16	ndhD	NADH dehydrogenase, ND4
rps18	protein CS18	ndhE	NADH dehydrogenase, ND4L
rps19	protein CS19	ndhF	NADH dehydrogenase, ND5

about 120 are present in the chloroplast genome. All other plastid proteins are expressed from nuclear genes (Gruissem, 1989c). Thus, there must be a significant flow of information between these two separate genetic compartments in the cell (Gruissem, 1989a).

In plant cells, protein synthesis occurs in organelles, chloroplasts and mitochondria, as well as in the cytoplasm. The machinery involved in transcription and translation of chloroplast genes is chloroplast specific and differs from that used in the expression of nuclear genes (Steinmetz and Weil, 1989). Translation of chloroplast genes requires the synthesis of both rRNAs and tRNAs, first, before proteins translated from 100-150 different gene transcripts (Sugiura, 1989a) can be obtained.

C. Ribosomal RNA genes

Chloroplasts contain the 70S class of ribosomes and share a considerable number of features with the prokaryotic ribosomes and are distinct from those found in the plant cell cytoplasm (Steinmetz and Weil, 1989). Chloroplast ribosomes are found free in the stroma, and associated with thylakoids (Steinmetz and Weil, 1989). Ribosomes are approximately 2/3 RNA by weight, and rRNA has a considerable structural role in ribosomes (Sibbald, 1988). In tobacco, approximately one third of the ribosomal proteins contained

in the ribosomes are encoded by chloroplast DNA (Shinozaki et al., 1986).

The 23S, 5S and 4.5S rRNAs are associated with the 50S subunit and the 16S rRNA is associated with the 30S subunit. All chloroplast *rrn* (ribosomal RNA) operons in higher plants show the typical organization of the prokaryotic *rrn* operons with the individual ribosomal RNA genes linked (Sugiura, 1989a) and transcribed in the order 16S-23S-4.5S (in higher plants only (Bowman and Dyer, 1979)) -5S rRNA. The rDNAs in most higher plants and *C. reinhardtii* are located in IRs and therefore present in two copies per genome. The exceptions to this organization include broad bean and pea in which there is only one copy of each of the rDNAs and *Euglena* (Z strain) for which there are three copies of the rDNA cluster arranged tandemly and a fourth copy of the 16S rRNA gene (Gray and Hallick, 1978; Jenni and Stutz, 1979). The *C. reinhardtii* rDNA cluster consists of 16S-7S-3S-23S-5S genes. An 888 bp intron encoding a 489 bp open reading frame, related to mitochondrial maturases, is found within the 23S rRNA gene. The 23S rRNA is preceded by two small rDNAs (3S and 7S) (Rochaix and Malnoe, 1978). Interestingly, the *Chlorella ellipsoidea* rDNA cluster is split into two back-to-back operons. The first is 16S rRNA-trnI(GAU), the second, trnA(UGC)-23S rRNA-5S rRNA (Yamada and Shimaji, 1986). The trnA and trnI genes contain no introns, while the 23S rDNA contains an intron of 243bp (Yamada and

Shimaji, 1987).

Until recently, evidence for the transcription of the 4.5S and 5S rRNAs within the same primary transcript as 16S and 23S was questionable. It was concluded from S1 nuclease mapping and primer extension analysis of rRNA processing intermediates and mature rRNA species from maize chloroplasts that these rRNAs are part of the polycistronic transcript and not transcribed from separate promoters (Strittmatter and Kossel, 1984; Gruissem, 1989b). The spacer region between the 16S and 23S rRNA genes ranges from 1.6 to 2.4 kb in all higher plants and includes the two intron-containing tRNA genes, tRNA^{His} and tRNA^{Ala} (reviewed in Gruissem, 1989b). In *C. reinhardtii* and *Euglena*, the same tRNA genes are located in the 16S-23S spacer region, but they do not contain introns (reviewed in Gruissem, 1989b; Graf *et al.*, 1980; Rochaix, 1981). *E. coli* rrn operons also contain the tRNA genes trnA and trnI (Delaney and Cattolico, 1989).

rRNA gene expression and processing

The regulation of transcription of the rrn operons in chloroplasts is reminiscent of that of prokaryotic ribosomal RNA operons. Several chloroplast genes are organized into polycistronic transcription units (Gruissem, 1989b). Conserved DNA sequences have been found in the 5' region of the chloroplast 16S rRNA gene which can be folded into three

stem-loop structures, termed H1, H2, and H3 (Briat et al., 1983). Depending on the loops formed between the regions, the RNA polymerase may initiate transcription of the operon. Chloroplast RNA polymerase recognizes specific upstream promoter regions and for a number of genes such as that of the *rbcL* gene, for example, the sequences of these promoter regions are highly conserved in different higher plants (Gruissem et al., 1987). Two sequence elements *ctp1* and *ctp2* within these conserved regions are critical for promoter function and their organization is similar to that of the *E. coli* consensus promoter (reviewed in Gruissem, 1989b). Differential strengths of promoters are considered first controlling steps which establish basal transcription activities of individual transcription units (Deng et al., 1989). Also, two spinach chloroplast tRNA genes, *trnR1* and *trnS1*, were shown to require no upstream promoter elements for transcription (Gruissem et al., 1986). Promoter sequences internal to nuclear tRNA genes transcribed by RNA polymerase III have been documented (Gruissem, 1989b).

In higher plants, chloroplast genes tend to be multicistronic, the ends of the genes being demarcated by inverted repeats that are apparently involved in RNA stabilization and processing of the mature RNA (Stern et al., 1989; Stern and Gruissem, 1989). The abundance of ribosomal RNA in plastids correlates with a high rate of transcription which may result from selected transcription

of the rDNA transcription unit (Mullet and Klein, 1987).

Posttranscriptional processing of polycistronic and intron-containing RNAs most likely represents an important regulatory step in the control of chloroplast gene expression (Gruissem, 1989b). The mechanism of rRNA processing has been studied in detail mostly in *E. coli* and other prokaryotes, but little information is available for organellar processing. Processing of a precursor rRNA molecule involves multiple cleavage reactions at the 5' and 3' ends of each species of rRNA. The first cleavage that occurs in *E. coli* separates the 16S and 23S rRNAs from each other, the 5S rRNA and tRNAs, and is mediated by RNase III. The ensuing maturation of the 16S, 23S and 5S RNAs is accomplished by several different enzyme activities that have been partially purified (reviewed in Gruissem, 1989b) in chloroplasts. Chloroplast RNA processing appears to depend on enzyme activities distinct from those present in prokaryotes and is likely to be more complex, probably involving proteins for stability (Stern *et al.*, 1989; Stern and Gruissem, 1989). Polycistronic RNAs ensure the same ratio of synthesized RNAs.

Ribosomes are assembled in the chloroplast using rRNA species and ribosomal proteins from this organelle, in addition to ribosomal proteins imported from the cytoplasm (Steinmetz and Weil, 1989). In liverwort, for example, it is interesting that most of the ribosomal proteins in the

30S and 50S subunits encoded in the chloroplast genome are important in the initial stages of ribosomal protein-rRNA assembly (Wittmann, 1983); these ribosomal proteins are called the primary binding proteins (Noller *et al.*, 1987). These results indicate that to some extent, the expression of the chloroplast genome is co-ordinated with that of nuclear genome for the biogenesis and assembly of ribosomes (Fukuzawa *et al.*, 1988).

D. Transfer RNA genes

Chloroplast genomes are believed to encode all the tRNA species used in chloroplast protein synthesis. Thirty tRNA genes were identified by computer analysis of tobacco and rice chloroplast sequences (Shinozaki *et al.*, 1986a; Hiratsuka *et al.*, 1989), thirty one in liverwort (Ohyama *et al.*, 1986b) and subsequent analysis showed that all were expressed. The additional tRNA gene in liverwort is tRNA^{Arg} (CCG) (Ohyama *et al.*, 1986b). The tRNA genes are scattered over the chloroplast genome in higher plants, while in *Euglena*, most of the tRNA genes are clustered (Sugiura, 1989a). All possible codons are used in the sequences coding for polypeptides in chloroplasts. The minimum number of tRNA species required for translation of all codons is thought to be 32; however, no tRNAs recognizing codons CUU/C (Leu), CCU/C (Pro), GCU/C (Ala), and CGC/A/G (Arg) have been found. If the "two out of three" mechanism operates in the

chloroplast, these 30 tRNAs are likely to be sufficient to read all codons in the chloroplast system (Shinozaki et al., 1986a).

Processing of tRNAs

Six tRNA genes in higher plants harbor long single introns which range in size from 0.5-2.5 kb. Introns in the trnL gene can be folded into a Group I intron secondary structure whereas those of trnA and trnI can be folded into a Group II intron structure (reviewed in Gruissem, 1989b). This was first discovered in the maize trnA and trnI genes which are located in the spacer region separating the 16S and 23S rDNAs (Koch et al., 1981). In contrast, the tRNA genes trnA and trnI of the golden-brown alga Olisthodiscus luteus have no introns and are separated from each other in the 23S-16S rDNA spacer by only 3 bp (Delaney and Cattolico, 1989).

The mechanism for tRNA processing in chloroplasts is suspected to be analogous to that described for E. coli (reviewed in King et al., 1986). Although a direct mechanism for chloroplast tRNA processing has not been determined, an RNase P-like activity in tobacco produced the mature 5' ends of the chloroplast tRNAs in vitro (Yamaguchi-Shinozaki et al., 1987) suggesting similarity of the process to that of E. coli. The 3' -CCA is not encoded in the tRNA genes (Gruissem, 1989b) but is required for aminoacylation;

therefore, it must be added post-transcriptionally to the processed tRNAs by tRNA nucleotidyltransferase. Such a transferase activity has also been detected in a spinach chloroplast extract (Greenberg *et al.*, 1984). After transcription of a tRNA sequence, specific nucleotides within the tRNA are modified by substitutions of additional groups (Gruissem, 1989b). Modified bases have also been identified in chloroplast tRNAs (Greenberg *et al.*, 1984), but the enzymes catalyzing these reactions have not been isolated. All the tRNAs can form the cloverleaf structure and show closer homology to the bacterial tRNAs than to the corresponding eukaryotic cytoplasmic tRNAs (Sugiura, 1989a).

II. USE OF RIBOSOMAL RNA AS MODELS FOR EVOLUTIONARY STUDIES

Phylogenetic analyses have been traditionally performed on "phenotypical" data, ie. morphological, chemical, metabolic or behavioral, and have given rise to the discipline of numerical taxonomy in order to weigh, compare, and rationalize these data (reviewed in Cedergren *et al.*, 1988). With the advent of "genotypical" data, new types of databases were made available, ie. protein and nucleic acid sequences. Comparisons between phenotype-derived and sequence based phylogenies have drawn many similarities (Cedergren *et al.*, 1988).

A. Primary sequence of rRNAs

The rRNA structure has been strongly conserved throughout evolution because of its fundamental importance in protein synthesis (Yamada and Shimaji, 1987). Comparisons of ribosomal RNA sequences from diverse organisms provide insight into ancient evolutionary relationships. Furthermore, comparative sequence analysis is a powerful approach toward understanding rRNA structure and mechanisms of the translation process (Chao et al., 1984).

Sequence data are preferable to other molecular methods for assessing evolutionary relatedness because they allow quantitative interpretation and they form a growing data base for subsequent reference. In this respect computers play an invaluable role in the analysis of the sequences accumulated. The 16S-like and 23S-like rRNAs contain a sufficient number of residues which change at a rate proportional to the evolutionary distance considered (Pace et al., 1986) and are sufficiently conservative in structure and function that they can be used to demonstrate the most ancient relationships among all organisms. The 5S rRNAs and tRNAs also are conservative in structure, but they contain too few nucleotides to establish the more distant relationships accurately.

Comparisons of rRNA sequences have provided evidence of three lines of descent: archaebacteria, eubacteria, and the eukaryotic nucleus (Gray, 1989a). Phylogenetic trees

generated from the small subunit and large subunit rRNA species (such as presented in Figure 2, generated by a systematic statistical method called the "bootstrap method" (Cedergren et al., 1988)) have demonstrated that nuclear and organellar genomes are derived from different lineages (Figure 2). Organellar rRNA sequences are not only prokaryotic in character, but are of direct eubacterial ancestry. Moreover, rRNA phylogenetic trees show that chloroplasts are most closely related to cyanobacteria (Tomioaka and Sugiura, 1983), whereas mitochondria are specifically affiliated with the α -subdivision of the purple photosynthetic bacteria (Gray et al., 1989a; Yang et al., 1985; Gray et al., 1984). This proves that chloroplasts and mitochondria could not have simultaneously originated from within the eukaryotic cell (Gray, 1989a). These unrooted trees (where no single ancestor is defined) are in agreement with phylogenies established using 5S rRNAs (Hori et al., 1985; Hori and Osawa, 1987; Giovannoni et al., 1988; Van den Eynde et al., 1988).

B. Secondary structure of rRNAs

Phylogenies have initially been established using alignments of the primary sequences of ribosomal RNA genes from a number of organisms which make up the different lineages. In addition, these comparisons had brought about some interesting observations about the conservation of

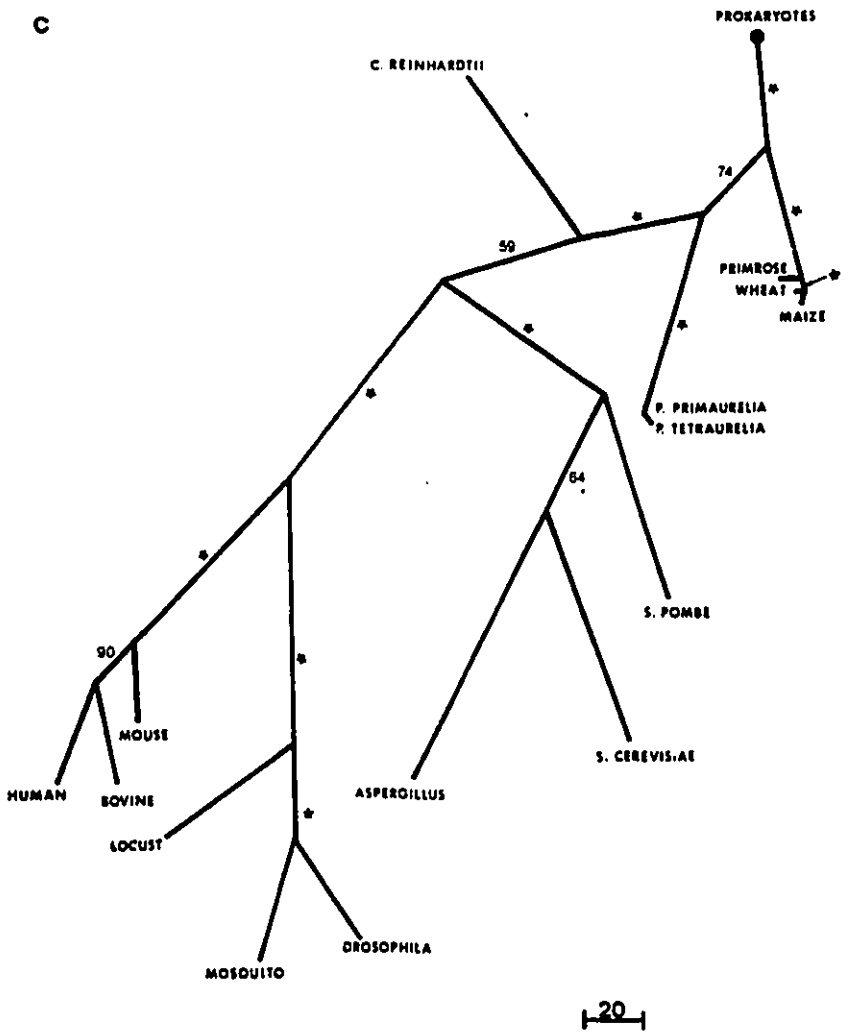
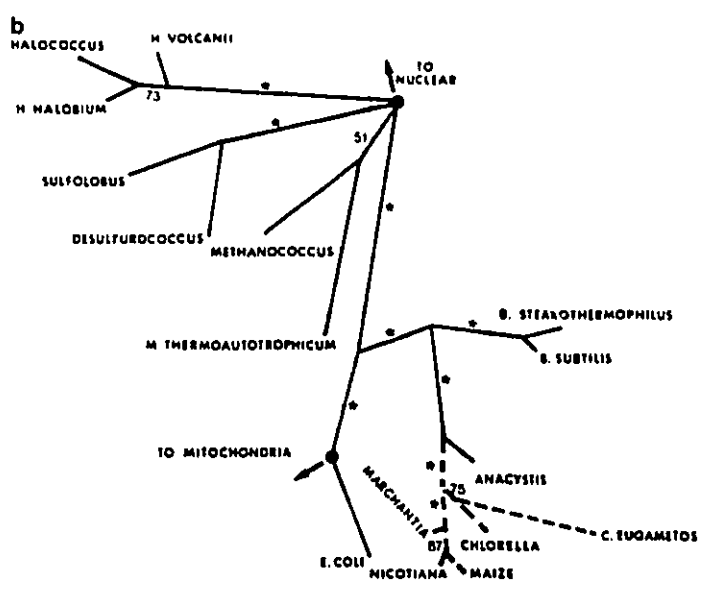
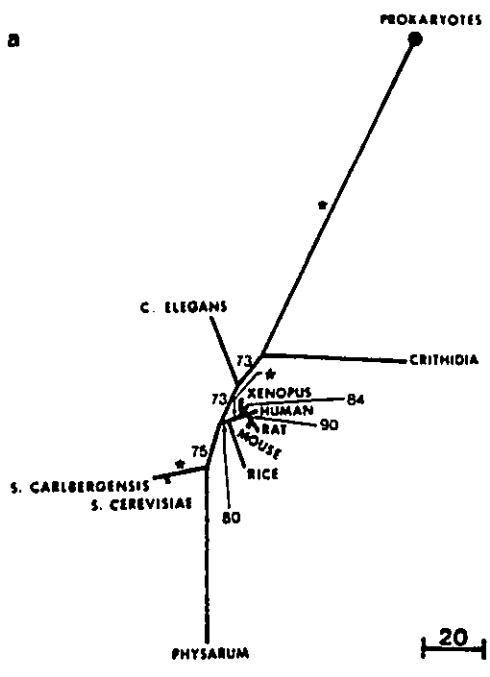
nucleotides at different positions throughout the ribosomal RNA molecules. The observation led to the determination of a model for the secondary structure for the SSU rRNA species and later for the LSU rRNA species.

Some 400 complete SSU-rRNA sequences (Gutell and Woese, 1990) have been published and most of these share a common core structure with variable regions extending from it (reviewed in Gutell *et al.*, 1987; Raué *et al.*, 1988), which is comprised of extensive intramolecular base pairing to form a series of helices, loops and bulges. The molecules are separated into three structural domains, each delimited by a long-range base pairing interaction, and a fourth minor domain which encompasses the 3'terminal 150 nucleotides of the molecule. The *E. coli* secondary structure for the 16S rRNA provides the basis of comparison between organisms and is presented in Figure 3. It was the first such structure to be determined (Woese *et al.*, 1983).

Structural analyses have also detected a secondary structure for the large subunit (LSU) ribosomal RNA species. Although the number of complete nucleotide sequences is about 70 (Gutell and Woese, 1990), studies similar to those for the SSU rRNAs have yielded comparable results. The LSU rRNA is about twice the length of the SSU rRNA which translates into a more complex secondary structure. Again, the *E. coli* structure was the first to be elaborated and is

Figure 2: PHYLOGENETIC TREES PREPARED FROM SMALL SUBUNIT
AND LARGE SUBUNIT RIBOSOMAL RNA MOLECULES

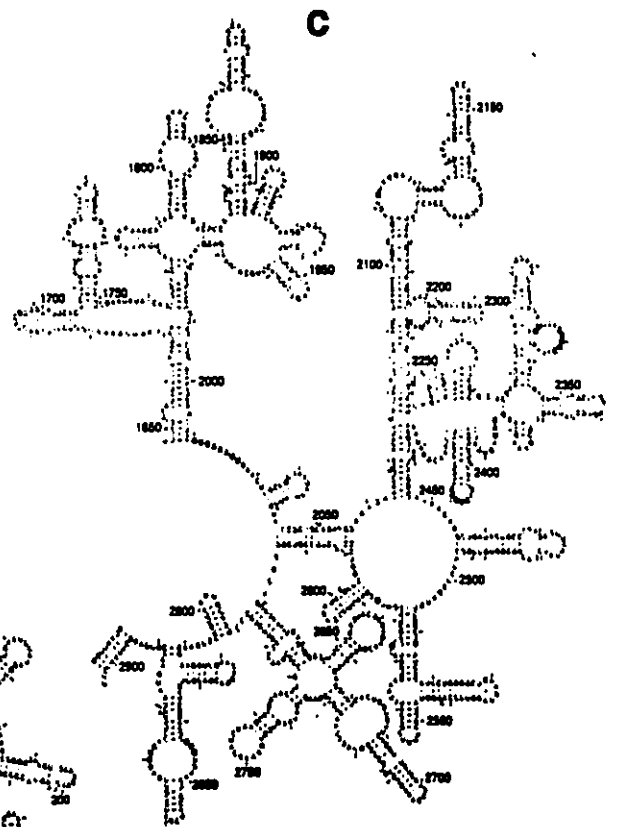
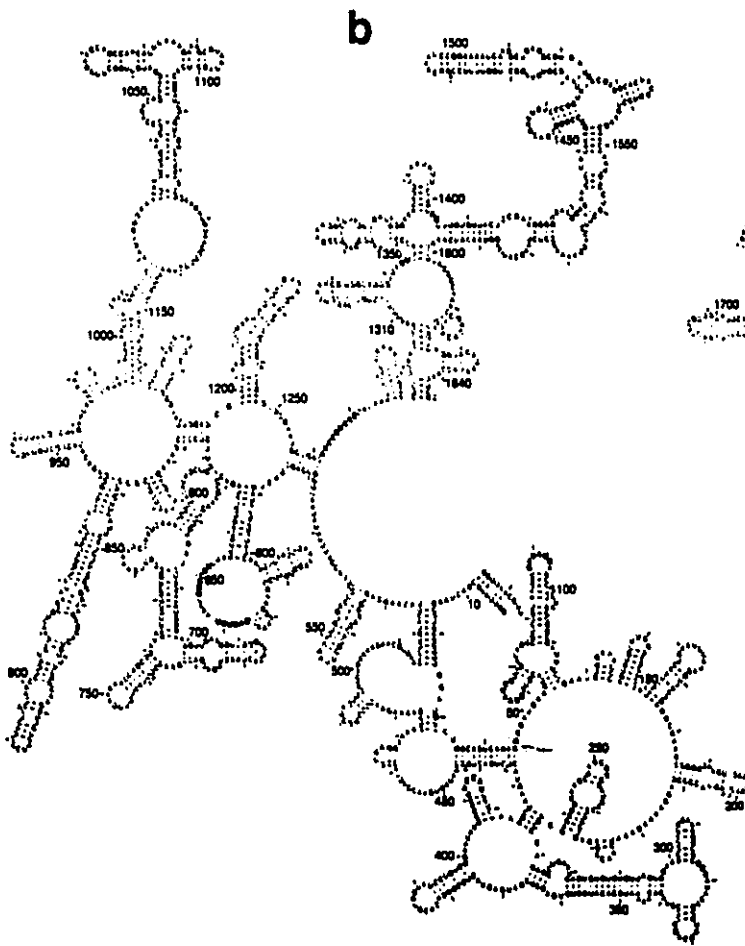
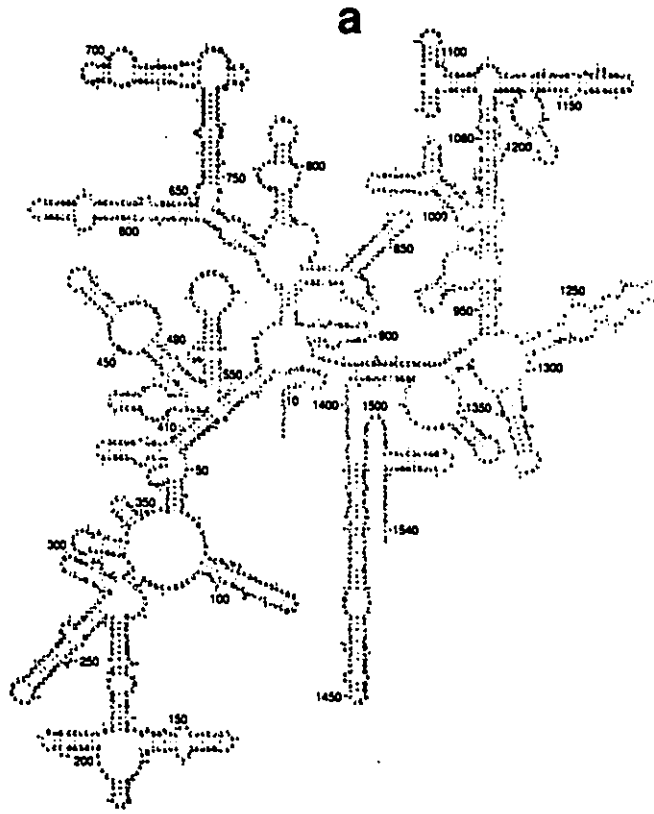
The trees are taken from Cedergren et al., 1988. A: the SSU tree; B: the LSU tree. In both trees, the lengths of the branches are proportional to the number of mutations (inferred mutational distance) indicated on the scales. Numerals on branches refer to the number of times a particular branch was found in 100 bootstrap samples. The star (*) symbol indicates that greater than 95% of bootstrap experiments contained this branch. a, Nucleocytoplasmic sequences; b archaeobacterial and eubacteria (chloroplast sequences are indicated by dashed lines); and c mitochondria. (...) A composite figure may be constructed by superimposing the appropriate solid circles that lead to the different lineages." (Cedergren et al., 1989; p. 105 and p. 106).



B

FIGURE 3: SECONDARY STRUCTURAL DIAGRAMS FOR THE E. coli 16S
AND 23S rRNA

The secondary structural diagrams for E. coli are presented. These diagrams are taken from Gutell and Woese, 1990 (p. 665) and represent the 16S rRNA (A) and the 23S rRNA with its 5' half (B) and 3' half (C). Every 10th position is marked and every 50th position is numbered, when possible.



presented in Figure 3. The structure can be divided into six structural domains. For both 16S and 23S rRNA secondary structures, conserved and variable regions are numbered according to the *E. coli* structure. Such secondary structures have provided insight on phylogeny in themselves since certain loops, or helices are conserved in one kingdom and not in another; these may be longer, shorter, absent or modified. Overall structures have corroborated already existent evidence that mitochondria are derived from the purple photosynthetic bacteria and chloroplasts from cyanobacteria.

In addition to their roles in the determination of phylogeny, secondary structures of ribosomal RNAs have also been studied to determine the roles ribosomal proteins serve in the ribosome, to elucidate the molecular mechanisms of assembly of large and small ribosomal subunits, to predict the three dimensional folding of the main domains of the rRNA structures and to identify sites of interaction of the ribosomal RNAs with tRNAs (reviewed in Noller *et al.*, 1987). Currently, efforts are placed upon defining probable sites of interaction of the secondary foldings of the rRNAs which would bring about their three dimensional structure (Gutell and Woese, 1990; Woese and Gutell, 1989; Gutell *et al.*, 1986).

It was deduced from phylogenetic studies that the chloroplast genome is a slowly evolving chromosome relative to other genomes; in a number of vascular plants, this genome is similar in size, gene content and gene order and gene sequences are highly conserved (Palmer, 1985).

III. ALDER AS A MODEL SYSTEM FOR SYMBIOSIS

Symbiotic nitrogen fixation plays a major role in soil nitrogen content thus providing natural fertilization to poor soils. Actinorhizal plants (non-legumes) which are mainly woody, perennial angiosperms (dicotyledonous shrubs or trees, with the exception of two species of *Datisca*) participate in symbiotic interactions with an actinomycete of the genus *Frankia*. Actinorhizal plants are members of 220 plant species belonging to 8 families and 23 genera (Dawson, 1986).

Due to their nitrogen fixing associations, these trees and shrubs are important in natural ecosystems and have potential value in forestry and land reclamation (Newcomb and Wood, 1987). They are found in a variety of environments, usually in poor soils and they are pioneering colonizers during the early stages of plant succession following disturbances such as fire, landslides, volcanic eruptions, flooding. Red alders (*Alnus rubra*) were found colonizing mud flows covered with ash resulting from the

eruption of Mount St. Helens in Washington state, USA; actinorhizal trees were planted to stabilize land disturbed by the construction of massive hydro electric power facilities in Quebec in the James Bay area; Alnus rubra, can be planted as a nurse crop for the lumber species Douglas fir (Pseudotsuga taxifolia) which uses the fixed nitrogen released by decaying alder leaf litter; Casuarina, distributed in tropical regions, is the world's best firewood in terms of kcal/unit mass (Dawson, 1986; Elmerich, 1984; Newcomb and Wood, 1987; Normand and Lalonde, 1986; Torrey, 1983; White and Williams, 1985; Wheeler et al., 1986). Actinorhizal plants also have a role to play in soil stabilization, increased soil nitrogen fertility, wood and fibre production, improved growth of associated timber, food, and forage crops (Dawson, 1986). The symbiosis between these plants and the microsymbiont warrants further investigation as they have many diverse uses in silviculture and agriculture.

Actinorhizal plants have evolved specialized root structures called nodules, which result from root tissue differentiation following microsymbiont invasion of the root hairs. The invasion process follows similar steps in legumes and in non-legumes (Newcomb and Wood, 1987), with minor variations. Expression of specific plant genes, nodulin genes, are essential for the structure and metabolism of the nodule. No legumes or monocotyledonous

plants are known to form nitrogen fixing symbiosis with Frankia (Newcomb and Wood, 1987).

The fixed nitrogen from actinorhizae is made available to the ecosystem on an extended basis through leaf fall, nodular decomposition, and plant death (Torrey, 1978). Nitrogen input depends on the frequency of nodulation, as well as the rate of fixation (White and Williams, 1985) and estimates of fixed nitrogen range depending on a number of variables. In ecosystems containing Alnus the annual nitrogen accretion was estimated to be up to 320 kg/hectare nitrogen for young stands of red alder in Oregon, USA (Dawson, 1986).

The presence of root nodules, which chemically reduce (fix) atmospheric nitrogen (molecular nitrogen) overcomes deficiencies of ammonium and nitrate in the soil and greatly aids plant growth (Dawson, 1983). Unlike alders, birch (Betula papyrifera) does not engage in symbiotic nitrogen fixation, although it is taxonomically related to alder. Therefore, birch may become a model system to study symbiotic associations of actinorhizal plants with Frankia species by transfer of plant symbiotic genes from alder to birch (Mackay et al., 1988).

The inheritance pattern for alder chloroplasts is not known, but is assumed to be maternal as is the case for most angiosperms. Paternal inheritance has been observed for this organelle in most gymnosperms (McCown and Ellis, 1989;

Palmer et al., 1988a). Inheritance of this organelle could be investigated if a breeding program for alders were to be established. At that time, chloroplast DNA of both parents and offspring could be isolated and restriction fragment length polymorphisms could be identified (Neale et al., 1989).

In summary, the alder symbiosis with the actinomycete Frankia merits further investigation. To date, much of the work done has been at the level of characterization of the bacterial genes (Normand et al., 1983; Simonet et al., 1986; Normand and Lalonde, 1986; Normand et al., 1988; Normand and Bousquet, 1989), but little is known about the alder genes: nuclear or organellar.

IV. OBJECTIVES OF THIS WORK

This project was initially focused on the isolation of nodulin genes, that is, plant genes which are involved in nitrogen fixation and metabolism of the nitrogenous compounds in the nodule tissue of the alder Alnus incana. The cDNAs used in this work namely pAnod2, pAnod4 and pAnod5 were thought to be nodule-specific as patterns of mRNA expression were reminiscent of those observed for nodulin genes: pAnod2 had a pattern which was that of early nodulins (Govers et al., 1987), whereas pAnod4 and pAnod5 showed that of late nodulins (reviewed in Verma and

Delauney, 1988). The sizes of mRNA determined from Northern blot analysis of poly A+ RNA of nodule versus root tissue for each cDNA probe, were not those documented for ribosomal RNA genes. A genomic library using total DNA was then constructed and screened using the cDNA inserts as probes in order to elucidate the structure of the alder nodulin genes. From the initial analyses of the clones selected from the library, it was proposed that the genes homologous to pAnod4 and pAnod5 might be linked as different portions of one set of lambda clones hybridized to both probes. Upon complete sequencing of the cDNA clones and comparisons with data banks, it became obvious that the clones were ribosomal in nature. Searches with partial sequences of the cDNAs versus organellar and plant genes had previously been done and had not detected significant homology with any other gene. It was later found that the copy of the analysis program used was defective.

Homologous sequences detected for the cDNA clones which were used as probes were as follows: pAnod2 showed the strongest homology to nuclear large subunit ribosomal RNA genes; pAnod4, to bacterial large subunit ribosomal genes, and pAnod5, to bacterial small subunit ribosomal genes, both probably Frankia. A more detailed presentation of the data which lead to this change in direction is presented in Appendix 1.

As this project was well underway when the results of the second data bank searches with the cDNA insert sequences were known, the analysis of the alder lambda clones obtained was redirected in light of the findings.

The present work will describe the isolation of chloroplast ribosomal RNA genes from the alder *Alnus incana*, and their subsequent characterization with respect to gene order, linkage and sequence as compared to known chloroplast ribosomal RNA genes, to help locate this species on a phylogenetic tree. A secondary structure for the 16S rRNA species is presented superimposed on the maize chloroplast 16S rRNA for comparison.

It may be possible to develop probes specific for the detection of the chloroplast genome in other angiosperms by defining sequences which would show homology to chloroplast DNA only.

CHAPTER 2: MATERIALS AND METHODS

I. ISOLATION OF DNA AND RNA

A. DNA isolation

1. Isolation of genomic DNA from plants

DNA isolation was done as described in Hattori *et al.*, 1987, up to step 12. Typically 15-20 g of leaf tissue harvested from young trees grown as described in Section X for alder (*Alnus incana*), or collected locally for birch (*Betula papyrifera*) were processed. The yield of DNA recovered ranged between 200-500 μ g. For alder, another step was added to further purify the DNA. An anion-exchange resin (Qiagen^R, Qiagen Inc., Studio City, California) was used following the manufacturer's instructions.

2. Small scale DNA preparation from positive lambda phages

a) Preparation of the lysate

A small scale phage lysate was obtained by resuspending a single plaque in 5 ml of L broth* with 15 mM MgSO₄ in a glass 15 ml tube, to which was added 50 μ l of an overnight culture of WA802. The cultures were shaken vigorously at

* Indicates found either in the list of abbreviations or in the Buffer section XII.

37°C for 5 hours, at which time clearing of the culture was usually visible. When this was not observed, the culture was diluted 1:3 with warm broth containing 15 mM MgSO₄ and was left to shake for an additional hour. A few drops of chloroform were then added to the lysate, the tubes vortexed to lyse intact cells and centrifuged 15 minutes at 4°C at 8000 rpm to pellet debris. The supernatant was collected in sterile vials. The titre was routinely 5 X 10⁹ pfu/ml or greater.

b) Extraction of DNA

DNA extraction of small scale lysates was done according to H. Murialdo (M. Whiteley, personal communication). To a lysate of 2 ml per DNA extraction were added 20 µl of 1 mg/ml DNase and the mixture was incubated at 37°C for 1 hour. Following the addition of 0.3 ml of 5 M NaCl and 0.75 ml of 40% (w/v) PEG 8000 (cold), the tubes were incubated on ice for 1 hour. After centrifugation at 5000 rpm for 10 minutes at 4°C, the pellet was resuspended in 0.2 ml cold SM buffer* and vortexed. This was transferred to an Eppendorf tube and extracted twice, sequentially with 0.2 ml chloroform. To the aqueous phase were added 4 µl of 0.5M EDTA*, pH 8 and 6 µl of 10% SDS*, proteinase K (Boehringer Mannheim) to 100 µg/ml, and it was incubated 20 minutes at 65°C. The aqueous phase was extracted sequentially with 0.3 ml of phenol saturated with

Tris pH 8, phenol/chloroform and SEVAG. DNA was precipitated overnight at -20°C with the addition of NaCl to 0.3 M and 2.5 sample volumes of 95% ethanol. DNA was collected by a 5 minute centrifugation, followed by washes with 80% and 95% ethanol, respectively and resuspended in 100 μl of TE buffer pH 8. One microliter of 10 mg/ml RNase was added to the reaction mixtures during restriction to remove the RNA present in the preparation.

3. Large scale DNA preparations from lambda phage

a) Preparation of the lysate

Large scale lysates were prepared as follows: 2.5 X 10^6 phage (obtained from the small scale lysate preparation) were incubated with 1×10^9 cells (multiplicity of infection of 0.025) for 20 minutes at 37°C . Of this mixture 40% was used to inoculate 500 ml of prewarmed L broth supplemented with 15mM MgSO_4 . The cultures were shaken vigorously overnight at 37°C and 10 ml chloroform were added. The lysates were titred before continuing the isolation procedure. Typically, titres of 3×10^9 pfu/ml were achieved.

b) Extraction of DNA

The purification of lambda DNA was done according to Maniatis et al., 1982, except that cesium chloride was added

to 0.8g/ml and centrifugation was done in a Beckman 50 VTi rotor for 19 hours at 15°C at 44,000 rpm.

4. Isolation of restriction fragments from agarose gels

Restriction digestions were electrophoresed on a 0.8% agarose gel, with Lambda DNA digested with HindIII and ϕ X174 DNA digested with HaeIII as standards. Fragments were excised from stained gels, and electrophoresed through a second gel to ensure separation of the fragments. Fragments were isolated from the gel with GeneClean following the manufacturer's (BIO 101) instructions and the amount of DNA was estimated by staining with ethidium bromide using ϕ X174 DNA digested with HaeIII as control.

5. Isolation of plasmid and M13 phage DNAs on a small scale.

a) Plasmid DNA

Small scale preparation of plasmid DNA (pTZ18R, Pharmacia) was done by a modified alkaline lysis method of Birnboim and Doly, 1979 (Morelle, 1989). Typically, 2.5 % of the DNA obtained was sufficient to analyse on an agarose gel. The resulting DNA was used for restriction analyses and sequencing.

b) Phage DNA

i) DIGE (DIrect Gel Electrophoresis)

M13mp19 (Messing, 1983) phage supernatants were

electrophoresed on a 0.8% TBE* (Maniatis et al., 1982) agarose gel to verify approximate insert size and hybridization to the proper probes.

ii) Preparation of DNA

Small scale preparation of phage DNA was done by PEG/NaCl precipitation, according to the M13 cloning and sequencing handbook from Amersham, except that L broth was used for growth of bacteria instead of 2X TY. The phage DNA prepared was then used as single-strand template for sequencing.

B. RNA isolation

Alder RNA isolation

RNA was isolated according to the procedure of Logemann et al., 1987, with the following modifications. An extraction with phenol/chloroform and precipitation step with 1/10 volume of 3M sodium acetate pH 5.2, and 2.5 volumes of 95% ethanol were added at the end of the protocol. All labware and buffers which were autoclavable were sterilized for 50 minutes prior to their use. Typically 5 g of tissue were extracted to yield approximately 55 μ g of total RNA as determined by optical density readings at 260 nm.

II. SOUTHERN AND NORTHERN TRANSFERS, PLAQUE LIFTS

A. Southern transfers

1. Standard procedure

Briefly, the DNA (in an agarose gel, or dotted onto a Biotrans membrane) was denatured in 1.5M NaCl / 0.5M NaOH and neutralised in 3M sodium acetate as described in the ICN protocol handbook. When DNA was dotted onto a membrane (for control dots or plaque lifts), the membrane was placed DNA side up on blotting paper soaked in denaturation and neutralization solutions, and the membranes were left to air dry. For agarose gels, DNA was allowed to transfer by capillarity with 20X SSC' as the transfer buffer (Southern, 1975). Following transfer/ denaturation, the DNA was crosslinked to the membrane by UV' illumination (Black Ray lamp, Ultra-Violet Products, Inc., San Gabriel, CA) for 5 minutes.

2. Genomic transfers

Five micrograms of genomic DNA per sample were restricted with various restriction enzymes according to the manufacturers' instructions and electrophoresed on a 0.8% agarose gel in TAE' buffer (Maniatis et al., 1982). After staining with ethidium bromide and visualization by UV light, the gel was

pretreated with 0.2N HCl for 5 minutes and processed as described in 1.

3. Dot blot controls

Dot blot controls were added to every hybridization done in order to determine the sensitivity of the probe after an overnight or other appropriate exposure. Plasmids containing the control DNAs were digested with restriction endonucleases as per manufacturers' specifications. These were diluted in 10 X SSC, to 10, 1, 0.1, 0.01 and 0.001 ng/2 μ l and spotted on a Biotrans membrane, then air dried and prepared as Southern transfers.

B. Northern transfers

Five micrograms of RNA per sample were electrophoresed on 1.5% denaturing agarose gel containing formaldehyde as described in Maniatis et al., 1982. Lanes containing the E. coli rRNA markers were stained and the unstained portion of the gel containing the sample RNAs was used for transfer. Following a one hour wash of the gel in distilled water, the transfer was performed onto a Biotrans membrane overnight with 20 X SSC. The RNA was crosslinked to the membrane by UV treatment, and following transfer, the gel was stained as described in Maniatis et al., 1982 to verify complete transfer of the RNA.

III. PROBE PREPARATION

The DNA fragments were labelled with 32 -P dCTP by the random priming method (Feinberg and Volgelstein, 1984) using a kit from Amersham. The labelled fragments were then purified over a Sephadex G50M (Pharmacia) spun column as described in Maniatis *et al.*, 1982. The probe specific activity ranged from 4×10^6 - 1.2×10^9 counts per minute per μ g labelled insert.

IV. HYBRIDIZATIONS

The hybridization protocols outlined here are those of ICN Biotrans membranes (ICN Biomedicals, Inc.)

A. Hybridizations and washes

1. Southern transfers (Method A, of reference)

Membranes were prehybridized with 4 ml buffer/100cm² membrane, and hybridized with 2 ml/100cm² membrane.

Hybridization solution: 5X Denhardt's solution, 5X SSPE, 0.2% SDS, 500 μ g/ml single-stranded Herring sperm DNA.

Prehybridization was done for 1 hour and the hybridization overnight, both at 65°C.

Blots were washed 3 X 30 minutes at room temperature with 250 ml/100 cm² membrane of wash buffer (5mM sodium

phosphate pH 7, 1mM EDTA, 0.2% SDS). The filters were then exposed to Kodak XAR-5 film, with the aid of Quanta III intensifying screens.

2. Northern transfers (Method B of reference)

Membranes were prehybridized with 4 ml buffer/100 cm² membrane and hybridized with 2 ml buffer/100 cm² membrane. Hybridization solution: 5X Denhardt's solution, 5X SSC, 50 mM sodium phosphate, pH 7, 0.1% SDS, 250 µg/ml single-stranded Herring sperm DNA and 50% (v/v) deionized formamide. Prehybridization was done at least one hour and hybridization overnight, both at 42°C.

After hybridization, transfers were first dipped in 2X SSC, 0.1% SDS, then washed four times in the same buffer with vigorous agitation for 5 minutes each at room temperature and finally washed in 0.1X SSC, 0.1% SDS at 250 ml/cm² twice for 15 minutes at 50°C. The membrane was then exposed as described in 1.

B. Rehybridizations

Transfers were stripped of probe as follows: (i) one hour at 65°C, in 50% formamide (non-deionized), 10mM sodium phosphate pH 7, at 100 ml/100 cm² (ii) fifteen minutes at room temperature in 2X SSC, 0.1% SDS at 250 ml/100 cm². It was often necessary to repeat this process before the probe was removed completely. Care was taken not to let the

membranes dry between rounds of hybridization, washing, and stripping. Stripped membranes were exposed on Kodak XAR-5 film for at least 48 hours to ensure that the probe had been removed properly.

V. ALDER GENOMIC LIBRARY CONSTRUCTION

A. Isolation of MboI fragments for cloning

Alder DNA (30 μg) was partially digested with MboI (0.07 U/ μg) and electrophoresed in a 0.5% LMT (Low Melting Temperature) agarose gel to allow for separation of the large fragments. The gels were run at 4 °C in precooled TAE buffer, overnight. Lanes containing the markers were stained with ethidium bromide and visualised under UV light to identify fragments between 12-23 kb. These were isolated from the agarose gel with GeneClean, in accordance to the manufacturer's protocol, using 3 ml of NaI per g agarose. The isolated fragments were then analysed by agarose gel electrophoresis to test for size of insert, amount and ability to ligate. Approximately 0.9 μg of insert DNA were purified in this way. The insert DNA was then precipitated using 0.5 sample volume of 7.5 M ammonium acetate and 2 volumes of 95% ethanol.

B. Ligation of MboI fragments to LambdaGEM-11 vector

The precipitated fragments were pelleted, washed with

80% and 95% ethanol, dried at room temperature and resuspended in water in the presence of LambdaGEM-11 BamHI arms (Promega). The DNA concentration in the ligation mixture was about at 250 $\mu\text{g/ml}$ (Whittaker *et al.*, 1988). After the addition of ligation buffer* and T₄ DNA ligase, samples were incubated at 4°C and warmed slowly to 18°C overnight.

VI. LIBRARY SCREENING

A. Plating of library

Ligation mixtures were packaged *in vitro* using the Packagene Lambda DNA Packaging System from Promega according to the manufacturer's protocol. Part of the library (approximately 3×10^5 phage) was plated for amplification as recommended by Stratagene (Gigapack Gold, La Jolla, CA). The remainder of the library was plated using the host WA802 (Wood, 1966) on a 24 X 24 cm² Nunc bioassay dish. The plates were incubated inverted at 37°C for 8 hours. At this time plaques did not come into contact with one another. Plates were stored at 4°C. Approximately, 5×10^5 phage were plated for screening.

That portion of the library which was amplified was done so according to the Stratagene protocol as follows: 3×10^4 phage were incubated with 3×10^8 cells at 37°C for 15 minutes, and plated on 150mm Petri dishes with 6.5 ml LC top

agar*. The plates were then overlaid with 10 ml SM buffer (Maniatis et al., 1982) and left overnight at 4°C with gentle shaking. The next day, the SM buffer containing the phage was removed using a sterile pipette, and the plates rinsed with 2 ml buffer. The resulting pool of phage was titred and stored at 4°C under 0.3% chloroform. Typical titres were 5.8×10^9 pfu/ml.

B. Screening of genomic library

The 5×10^5 phage on the Nunc plate were transferred onto an autoclaved BioTrans membrane and treated as described in the Southern transfer section II.A.1. The membrane was hybridized as described for Southern transfers section IV.A.1, sequentially with three DNA inserts obtained from T. Gleeson. Following hybridization, membranes were washed as described, before overnight exposure. The membrane was stripped between rounds of hybridization with the three probes, using buffers described in (IV.B).

C. Plaque purification

Positive plaques (regions of $\sim 1 \text{ mm}^2$) were picked from the plate and stored in SM buffer. Dilutions were made and 100 μl aliquots of 10^{-2} and 10^{-4} dilutions were plated onto 100mm Petri dish. Those plates with a manageable number of plaques ($\sim 250-350$) were lifted onto BioTrans membranes and hybridized as previously described.

D. Storage of the library

A portion of the amplified library (1 ml) was stored for posterity in screw cap vials at -80°C after the addition of DMSO to 7% (v/v).

VII. SUBCLONING INTO pTZ18R AND M13mp19 VECTORS

A. Vector preparations

DNAs were digested with the appropriate restriction enzymes as per the manufacturers' recommendation, and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim). EDTA and SDS were added to 20 mM and 0.5% respectively and the reaction heated at 65°C for 10 minutes. The vector was purified with GeneClean and tested prior to its use.

B. Ligations and transformations

Ligation mixtures containing vector and fragment to a concentration of 4-5 $\mu\text{g}/\text{ml}$ were incubated overnight at 4°C and allowed to warm slowly to 18°C . One ng of each ligation was introduced into competent *E. coli* JM101 cells prepared for transformation by the method of Hanahan, 1985. A control of 10 pg for uncut M13mp19 and 1 ng for uncut pTZ18R was always added to verify transformation efficiency which

was usually 3×10^7 plaques/ μg vector for M13 and 6×10^6 colonies/ μg vector for pTZ18R.

VIII. DNA SEQUENCING

A. Preparation of templates

1. Template orientation

Complementation tests (C-tests) were performed with the phage DNA to determine the relative orientation of the clones to help in sequencing. Briefly, $1.3 \mu\text{l}$ of each DNA tested were added to $15.4 \mu\text{l}$ of water and $2 \mu\text{l}$ of 20 X SSC, and incubated at 65°C for 90 minutes. The samples were electrophoresed on a 0.8% TBE gel with $3 \mu\text{l}$ of M13 loading buffer* and $1 \mu\text{l}$ of 2% SDS. Some gels were subjected to Southern transfer analysis to confirm identity of the clones.

2. Dale deletions

Phage templates used were prepared as described in section I.A.5.b.ii. The reactions for (1) linearization of circular single-stranded DNA, (2) exonuclease digestion, (3) tailing reaction, (4) ligation were done as described in the Procedure Manual for IBI System RDS: Rapid Deletion Subcloning. There were some minor modifications. In the first step, $10 \mu\text{l}$ of the phage template was used ($\sim 4 \mu\text{g}$) and

the RD-22mer was annealed for 10 minutes at 65°C then allowed to cool to ~42°C slowly, before adding the EcoRI. The amount of T, DNA polymerase added in the second step was sufficient to delete 187 bp at 2.5 minute intervals. Other steps were executed as described. On average, 40 deletion clones which spanned the 2.5 kb fragments that were to be sequenced were obtained. Phage supernatants were subjected to DIGE (I.A.5.b.i), and the gel was blotted to locate those clones that hybridized to the original DNA probes.

B. Sequencing reactions

Sequencing reactions were done according to the dideoxy chain termination method (Sanger et al., 1977), using the USB Sequenase Kit (United States Biochemicals Corporation). Seven microliters of the prepared M13 phage template contained sufficient, clean DNA to be used for sequencing. α ³²S-dATP (NEN or Amersham) was used to label the DNA. Routinely, 350 nucleotides could be read from two loadings. Double-stranded sequencing was also done from pTZ18R subclones, using the procedure outlined in the Sequenase protocol. dGTP termination mixes were used for the majority of the sequencing reactions, whereas dITP termination mixes were used only for a few clones that had some compression problems. The reactions were then electrophoresed on a 7% acrylamide vertical gel in 0.5X TBE buffer. Two runs were executed; the first for 3 hours, the second for 1.5 hours.

The gels were run at 2100 V, 192 mA and 55 Watts. Gels were lifted onto 3MM blotting paper and dried at 80°C under vacuum. Exposure was done on 3M Medical Imagery Film for 24-48 hours.

IX. DATA ANALYSIS

The Sizefrag program (Schaffer, 1981) was used for some of the size determinations of restriction fragments.

Sequence entry, and DNA sequence manipulation was done with the PCGENE software package from Intelligenetics. Sequence comparisons and data bank searches were performed using the MicroGenie Sequence Analysis Program Release 6.0 (Beckmann Instruments, Inc., Palo Alto, CA) and IBI package which uses the Pustell Sequence Analysis Program version 2.02 (Copyright 1987-89).

X. GROWTH AND MAINTENANCE OF ALDERS

A. Growing Frankia ACN 14a

Frankia strain ACN 14a was obtained from M. Lalonde, Université Laval, Québec. QmodB culture medium (Normand et al., 1983) was used for growth of Frankia. The media was prepared, autoclaved and incubated at 30°C for at least three days to ensure sterility as this actinomycete grows slowly.

From an original inoculum, 10 ml of cells were transferred into 500 ml of fresh medium. Cultures were allowed to grow 33 days to obtain sufficient cell density to inoculate alder sprouts. It was found that the best growth occurred when the cultures were left standing with occasional swirling to provide aeration.

B. Sprouting alders (Alnus incana)

Seed from Alnus incana was purchased from F.W. Schumacher, Sandwich Mass., USA. Briefly, seeds were surface sterilized with 20% Javex and rinsed with distilled water. The seeds were spread evenly over wet vermiculite and covered with a thin layer of dry vermiculite in plastic trays. The sprouts were set at 20°C under growth lights with an 18 hour light/6 hour dark schedule. Sprouts were watered by spraying.

C. Planting and inoculation of alders

The sprouts were transplanted to 10 cm X 10 cm pots in Perlite watered in half-strength nitrogen-free watering solution (Cutting and Schulman, 1969). Alders were inoculated with Frankia (inoculum was passed through an 18 gauge needle prior to use to form a homogeneous suspension), by the addition of 40 ml per pot at the base of the plants. Transplanted sprouts received one addition of NH_4NO_3 to a final concentration of 1mM in order to increase plant

survival. Control plants received watering solution supplemented with 1mM NH_4NO_3 as nitrogen source. Plants were grown at 20-25°C with the same light schedule, under VitaLites (Duro-Test), for 14 weeks.

D. Collecting tissues

Leaf tissue was collected from the plants, frozen in liquid nitrogen, and stored at -80°C.

Root and nodule tissue was collected at 14 weeks (from the time of inoculation) from control plants or inoculated plants, after the Perlite was removed from the roots by a quick rinse in distilled water. The tissue was treated as described for leaf tissue. Nodules were hand-picked individually with forceps and dropped into liquid nitrogen. Tissue was stored for a maximum of 30 days.

XI. BACTERIAL STRAINS AND RELEVANT PHENOTYPES

A. For titre of genomic libraries prepared

LE392	McrA ⁻ , McrB ⁺ , supE, supF, hsdR514 (rk ⁻ , mk ⁺)
MB406	?mcr, supE, hsdR (rk ⁻ , mk ⁺)
NM538	McrA ⁻ , McrB ⁺ , hsdR514 (rk ⁻ mk ⁺), supF
NM539	(P2cox3) derivative of NM538
WA802	McrA ⁻ , McrB ⁻ , hsdR2 (rk ⁻ , mk ⁺), supE

B. For transformations with plasmid and M13

JM101	lac α , McrA ⁺ , McrB [?] , hsd (rk ⁺ , mk ⁺), supE, Δ (lac-proAB), [F', traD, proAB, lacI ^q lacZ Δ M15]
-------	---

References: Kaiser and Murray, 1985; Maniatis et al., 1982; Raleigh, 1987; Raleigh and Wilson, 1986; Raleigh et al., 1988; Whittaker et al., 1988.

XII. BUFFERS AND OTHER SOLUTIONS

Buffers and solutions indicated here are from Maniatis et al., 1982, unless otherwise stated.

- A. 7% acrylamide gel for sequencing: 33.6 g urea, 37 ml water, 14ml 40% bis/acrylamide (20:1), 4 ml 10X TBE, 56 μ l TEMED, 560 μ l ammonium persulfate (100 mg/ml).

- B. 50X Denhardt's solution: 1% (w/v) Ficoll (400 000 M. Wt.), 1% (w/v) polyvinylpyrrolidone (360 000 M. Wt.), 1% (w/v) bovine serum albumin, (Denhardt, 1966).
- C. L broth: (per l) 10 g tryptone, 5 g yeast extract, 5 g NaCl, pH 7.2.
- D. LC top agar: (per l) 10 g tryptone, 5 g yeast extract, 10 g NaCl, 0.72 g Tris base, 10 ml 1M MgSO₄, 5 ml 0.5M CaCl₂, pH 7.25 (7.2-7.3), 9 g agar.
- E. 10X ligase buffer: 400 mM Tris-HCl, pH7.5, 100 mM MgCl₂, 100 mM DTT, 10 mM ATP, 500 µg/ml BSA (Promega).
- F. M13 loading buffer: 50% (v/v) glycerol, 200 mM EDTA pH 8, bromophenol blue. (M13 cloning and sequencing handbook from Amersham).
- G. Phenol/chloroform: solid phenol was melted at 65°C until liquid, 8-hydroxyquinoline was added to 0.1% (w/v), and the phenol was equilibrated with 1M Tris pH 8.0 until the pH of the aqueous phase was ~ 7.9. An equal volume of 24:1 chloroform: isoamyl alcohol was added to the phenol.
- H. SEVAG: chloroform: isoamyl alcohol (24:1) v/v

- I. SM buffer: (per l) 5.8 g NaCl, 2 g MgSO₄.7H₂O, 50 ml of 1M Tris-HCl, pH 7.5, 5 ml of 2% gelatin.
- J. 20X SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0.
- K. 20X SSPE: 3.6 M NaCl, 0.2 M sodium phosphate, pH 8.3, 0.02M EDTA (ICN Biomedicals, Inc.).
- L. 20X TAE: (per l) 96.8 g Tris base, 32.8 g sodium acetate, 7.4 g disodium EDTA. pH 8.2 with the addition of glacial acetic acid.
- M. 10X TBE: (per l) 108 g Tris base, 55 g boric acid, 8.5 g disodium EDTA, pH 8.3.
- N. TE buffer: 10 mM Tris-HCl, pH 8, 1 mM EDTA, pH 8.

CHAPTER 3: RESULTS

I. SOURCE OF PROBES

The probes used in this work were derived from isolated inserts of three alder cDNA clones pAnod2, pAnod4 and pAnod5. Restriction enzymes which had no sites within the cDNA sequence (pAnod 2, EcoRI/BamHI 613 bp fragment; pAnod 4, EcoRI/SalI 1718 bp fragment; pAnod 5, EcoRI/BamHI 200 bp fragment) were used to remove the inserts cleanly from the vector (pGEM4B, Promega). The cDNAs had been obtained as a result of differential screening of a nodule cDNA library using end-labeled root and nodule RNA as probes (T. Gleeson, M.Sc. thesis). Initially these cDNAs had been thought to be nodule-specific clones (they were expressed in RNA extracted from nodule tissue only, according to the developmental expression of other nodule-specific clones), but upon sequence comparisons with data banks, they were found to be ribosomal in origin (see Appendix 1). The original names of the cDNAs were retained to simplify the analyses.

II. ISOLATION OF ALDER AND BIRCH DNAs

The method described for DNA isolation had been used in the past in our lab to purify DNA from a number of different plants. The age of the leaf tissue did not seem to influence

the yield of DNA but storage of the tissue at -80°C for periods greater than one month resulted in the isolation of degraded DNA. For the isolation of birch DNA, the method described in Hattori *et al.*, 1987 was followed until step 12 (resuspension in TE buffer) without further purification of the DNA. The yield of DNA recovered ranged between 200 - 500 μg for 15-20 g of birch or alder tissue.

Unlike birch, alder DNA isolated by this procedure was usually light to dark orange in colour and very viscous which created a problem for agarose gel electrophoresis and subsequent analyses. The DNA was too buoyant to stay in the wells of a horizontal agarose gel, even with loading buffer present and the migration of the DNA within the gel was retarded. The DNA itself however, like birch DNA, could be digested successfully with any restriction endonuclease tested (data not shown). A small sample of DNA (0.5-1 μg) could be electrophoresed if it was sufficiently diluted with water and loading buffer. This allowed us to verify digestion of the DNA as well as the integrity of the preparation. The colour was probably due to the presence of plant phenolics (C. Nozzolillo, personal communication) which seem to copurify with DNA and are very difficult to remove. A variety of procedures were attempted to correct the situation eg. high speed centrifugation, potassium acetate gradients to separate the DNA from the impurities, sodium acetate washes as prescribed for plant RNA

extractions (Logemann *et al.*, 1987) to remove polysaccharides (starch), but resulted in little success.

An anion-exchange resin linked to a silicagel base (Qiagen[®], Qiagen Inc., Studio City, California) was used to both remove colour and reduce viscosity of the DNA preparations. The recovery of DNA, however, was usually about 50% of the initial amount of DNA as determined by agarose gel electrophoresis and optical density measurements at 260 nm. It is noteworthy to mention that the initial absorbance reading was probably inaccurate due to the presence of the pigmented compounds in the DNA preparation, even if only a small amount was used and diluted for analysis. DNA purified in this manner was used to construct the alder genomic library and to prepare the Southern blots.

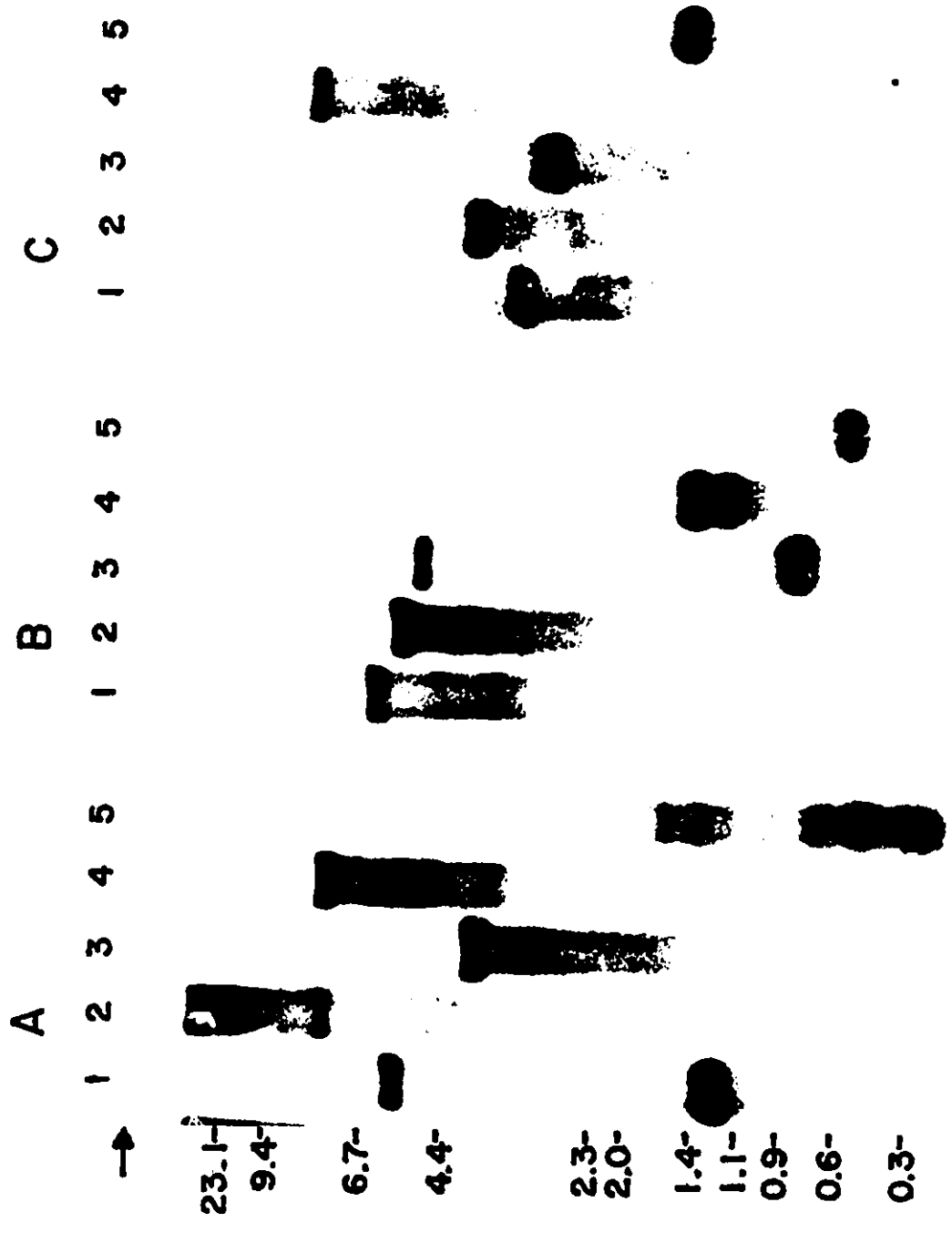
III. SOUTHERN TRANSFER ANALYSES

ALDER (*Alnus incana*) SOUTHERN TRANSFERS

The fragments hybridizing to each of the three cDNA probes are different and distinct (Figure 4). Initially, it was thought that the simple pattern and small number of restriction fragments suggested single copy (or low copy number) genes. For most enzymes, only one fragment hybridizes to the probes. The results show independence of the clones as each hybridized to different fragments.

FIGURE 4: HYBRIDIZATION OF ALDER DNA TO cDNA INSERT PROBES

A Southern transfer of alder genomic DNA was screened sequentially with cDNA inserts derived from A, pAnod2; B, pAnod4 and C, pAnod5. 5 μ g of DNA was digested with 1, BamHI; 2, BglIII; 3, EcoRI; 4, HindIII; 5, MspI. The arrow indicates the origin. Sizes of molecular weight standards (in kb) on the left of the panels are those of lambda DNA digested with HindIII and ϕ X174 DNA digested with HaeIII.



The sizes of the fragments hybridizing to the probes are found in Appendix 3; Table 5. These fragment sizes for selected restriction enzymes are the average of sizes obtained from four independent genomic blots.

IV. ALDER GENOMIC LIBRARY CONSTRUCTION

A. CHOICE OF RESTRICTION ENZYME FOR PARTIAL DIGESTION OF GENOMIC DNA

In order to obtain a random representative library, genomic DNA must be fractionated in a sequence-independent manner (Kaiser and Murray, 1985). The number of clones required to represent the alder genome was estimated using the soybean genome size for calculations as the alder genome size is not known. For a haploid genome size of 1×10^9 bp and a probability of 99.9% of representation, according to calculations described in Maniatis *et al.*, 1982 for a 17kb insert size, the number of recombinant lambda clones necessary would be 4.1×10^5 (2.7×10^5 clones would be necessary for a probability of 99%).

If DNA is fractionated following digestion with a restriction endonuclease which cuts relatively frequently as compared with the desired fragment size, then partial cleavage should produce a set of random overlapping

fragments. Sau3A and MboI are enzymes of choice in this respect as both recognize a tetranucleotide DNA sequence (GATC) which would occur, on average, every 256 (4²) base pairs in a random sequence DNA. MboI is insensitive to methylation at the 5MeC position whereas Sau3A is methylation sensitive at this position. This becomes important as plant DNAs are heavily methylated at CpNpG and CpG (Belanger and Hepburn, 1990). For the addition of the same number of enzyme units and constant amount of genomic DNA, MboI restricted alder DNA to a greater extent than did Sau3A (data not shown) suggesting methylation of the DNA. The use of Sau3A may have biased the library as all the available GATC sites may not have been accessible to be cleaved in a random fashion and DNA regions which were especially methylated would have been underrepresented. MboI fragments can be ligated into BamHI sites. Therefore MboI was chosen for library construction.

B. ISOLATION OF MboI FRAGMENTS FOR CLONING

In a preliminary experiment, alder DNA was digested with dilutions of MboI and 0.07 U/ μ g DNA of MboI was determined empirically to be the amount of enzyme required for partial restriction. The digest was then scaled up and electrophoresed in LMT (low melting temperature) agarose within an ordinary agarose frame for support. The gels were

run at 4°C in order to maintain the shape of the gel. The marker lanes and one test lane for the digest were stained with ethidium bromide and visualised under UV light to locate the fragments between 12-23 kb. DNA was isolated from the agarose as described in the Methods section of this work taking care to perform the purification steps gently since long fragments of DNA are easily sheared.

C. LIBRARY CONSTRUCTION

1. Choice of vector

The vector LambdaGEM-11 (Promega Corporation) has specific features making it the choice for the library construction. This vector is a derivative of EMBL3 (Frischauf *et al.*, 1983), has Spi phenotype selection available (Kaiser and Murray, 1985), SP6 and T7 promoters which allow direct sequencing and "chromosomal walking", and SfiI restriction sites located outside the promoter regions in asymmetrical positions to permit the rapid mapping of clones by using labeled linkers specific for SfiI sites in the cloning region at either end of the insert DNA (Promega Corporation).

2. Characterization of the library

Two alder genomic libraries were constructed in order to test reproducibility of the procedure. The titre of

phage was similar in both libraries (Table 2).

The average insert size for the alder genomic library was estimated by taking 25 plaques at random and isolating DNA on a small scale. These preparations were digested with XhoI to remove the inserts from the vector. The sizes ranged between 12-22 kb (data not shown), with the average size estimated to be 17 kb.

TABLE 2: Preparation of the alder genomic libraries; cloning summary

	<u>Library 27</u>	<u>Library 26</u>
ug vector LambdaGEM-11	3.6	2.4
ug insert alder genomic	0.9	0.6
recombinants on WA802	8.3×10^5	7.2×10^5
pfu/ml	8.3×10^5	7.2×10^5
pfu/ μ g vector	2.3×10^5	3.0×10^5

The vector background was tested by in vitro packaging and plating of uncut genomic DNA to 1.5 μ g vector. There were 9 plaques total, which represented 6 pfu/ μ g, a value below that claimed by Promega (20 pfu/ μ g; Promega, Certificate of Analysis April 12, 1987).

Given that 4.1×10^5 clones were calculated as necessary to represent the alder genome for insert sizes of

17 kb, and 8.3×10^5 clones were obtained as a result of the construction of the library (library # 27, for example), then the library would be representative if the initial assumption of genome size was accurate. The portion of the library (#27) (approximately 3.0×10^5 clones) that was not plated, but amplified, would be a sufficient number of clones to represent the alder genome at a 99% probability, if calculated as previously described.

D. CHOICE OF HOST STRAIN

The titre and platings of the initial genomic libraries constructed were performed using the strains NM538/NM539 as hosts. It was noticed that the titres were quite low (ie. in the order of 5×10^4 pfu/ml packaging extract) regardless of the amounts of phage DNA used. Other strains were also tested for titring the libraries ie. MB406, LE392 but the titre obtained varied greatly.

During the preliminary attempts to construct an alder genomic library, the effect of the M_{cr} phenotype was beginning to be reported in the literature. This Modified cytosine restriction is due to an E. coli system which restricts incoming DNA containing methylated cytosine residues (Raleigh and Wilson, 1986) and affects the yield of recombinants in specific libraries (Whittaker et al., 1988). As plant DNAs are highly methylated, the strain WA802 (E.

coli Genetic Stock Centre, Yale University, CGSC #5616) which has no mcr system was then tested. We observed a two fold increase in the number of recombinants titred on WA802 versus NM538, whereas Whittaker et al. saw an increase of up to tenfold with a cloning efficiency of $> 10^7$ pfu/ μ g human DNA (cf: Table 4 in Whittaker et al., 1988). This strain was therefore used for all other manipulations with bacteriophage lambda. Note that the strain WA802 was sent to Dr. W. Cheliak's laboratory at the Petawawa National Forestry Institute and used as host for a pine genomic library. Similar increases were obtained for the titre of the phage (L. de Verno, personal communication).

Overall, the titres using strain WA802 as host were higher than those using NM538, NM539, LE392 and MB406 as host (these last four strains had been provided with the lambda vector). Titres for library #26 were obtained on the five different strains mentioned, whereas for library #27, only three strains were tested (NM538, MB406 and WA802). Titres were highest on WA802, for both libraries ie. 7.2 and 8.3×10^5 pfu/ml respectively.

E. LIBRARY SCREENING

1. Initial screening

The three cDNA inserts were used to screen library #27. Figure 5 represents autoradiograms of the initial screening

and further plaque purifications using the pAnod4 insert as probe. For pAnod4 probe, 676 plaques from 500 000 plaques screened were positive (0.14% of total) and 10 were purified. After two rounds of purification, four of the plaques still hybridized to the probe (0.05% of total); the other six were false positives. Similar results were obtained for pAnod5 and pAnod2 where 8 and 7 clones, respectively, were first selected but 1 clone, each, purified true as defined by Southern transfer. As the frequency for the sequences is ~0.05% of the total clones in the gene bank, the sequence selected is either not single copy as proposed from total DNA alder Southern transfers (Figure 4), or this DNA is over represented in the bank.

2. Preliminary restriction analysis of isolated clones

Preliminary analysis of the positive clones was first performed by Southern transfer of small scale DNA preparation and hybridization to pAnod2, pAnod4 and pAnod5 inserts to verify that the fragment sizes observed on Southern transfers for a particular enzyme were those observed for the clones. DNA was isolated on a larger scale for subsequent analyses and subcloning of fragments. The sizes of insert of each selected clone were determined by XhoI digests and/or restriction mapping, and are presented below:

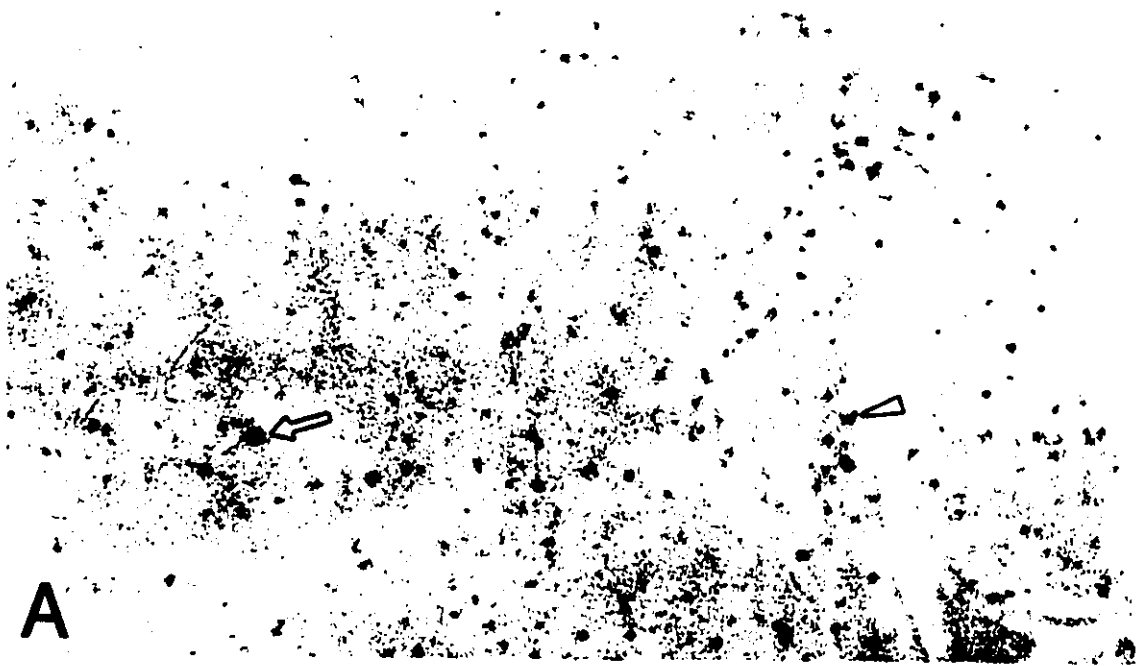
FIGURE 5: ALDER GENOMIC LIBRARY SCREENING

Autoradiograms of plaque lifts are represented to illustrate the initial alder genomic library #27 screening and subsequent plaque purifications using the pAnod4 insert as probe.

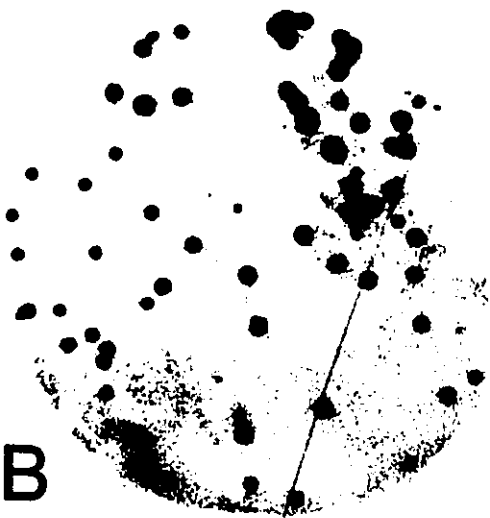
Panel A represents a portion of the autoradiogram for the initial screening of the library. The hybridizing spot at the end of the arrow contained a plaque that purified true; whereas the spot at the end of the arrowhead did not.

Panel B represents a first plaque purification of the clones contained in the plaques covered by the spot at the end of the arrow. 68 of the 372 plaques found on this plate (18%) were positive.

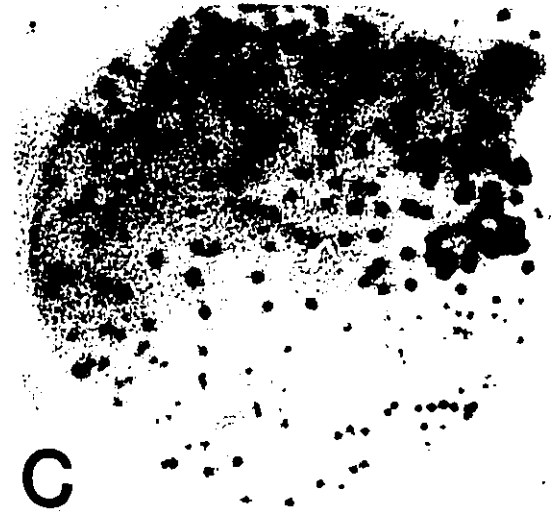
Panel C represents the second plaque purification step: all plaques hybridized to the probe. A positive plaque from this last purification was picked, replated and a plaque from this latter plate was used to do the small scale lysate.



A



B



C

<u>Clone selected</u>	<u>Size (kb)</u>	<u>Probe used</u>
λ Anod2-7	14.0	pAnod2
λ Anod4-1	18.6	pAnod4
λ Anod4-2	22.3	pAnod4
λ Anod4-4	19.2	pAnod4
λ Anod4-9	13.8	pAnod4
λ Anod5-5	13.4	pAnod5

Phage were named λ (from lambda genomic library) Δ (for alder) nod (for what was thought to be nodulin clones) # (for probe used for the screening) -# (for the isolate picked and purified from this screening).

Subsequent work was limited essentially to analysis of two clones hybridizing to pAnod4, namely λ Anod4-1 and λ Anod4-2. A clone hybridizing to pAnod5, λ Anod5-5 was partially analysed (see Appendix 2) and a clone hybridizing to pAnod2 (λ Anod2-7) was isolated but the analysis was not pursued.

V. RESTRICTION ANALYSIS OF CLONES PURSUED IN THIS WORK

Approximately 40 kb of alder sequences, cloned into λ Anod4-1 and λ Anod4-2, were mapped by restriction analysis. In addition to the more conventional methods for restriction site mapping, ie. double digests with combinations of restriction enzymes, the SfiI linker mapping system (Promega) was also tried to help in mapping the clones. It

involves asymmetric SfiI sites situated one in each lambda arm, which generate specific ends of insert DNA. These ends can be labeled by ligation of an oligonucleotide complementary to the right or left arm sites. A series of restriction enzyme dilutions allows for recognition of consecutive restriction sites. Some problems were encountered with the marker sent in the kits, and there was doubt as to the specificity of the right oligonucleotide. Only a small amount of the data gained by this method was used to order the BamHI fragments of clone λ Anod4-2. Therefore, the more conventional method was used for mapping the clones. The results obtained are summarized in Figure 6.

The maps for both cloned inserts of λ Anod4-1 (18.6 kb) and λ Anod4-2 (22.3 kb) are virtually identical over a defined region. Therefore, it was tentatively concluded that overlapping sections of DNA had been cloned and selected independently from the library. A 6.8 kb portion of the λ Anod4-1 clone was further analysed by sequencing (Figure 8), and some sequence information was obtained from selected subclones derived from λ Anod4-2 (data not shown) in order to confirm the restriction map assignments. The sizes of hybridizing fragments obtained for the digests of total alder DNA in Figure 4 are given in Table 5 (Appendix 3). In most cases, the size of fragments on the cloned insert DNA (Figure 6) hybridizing to the same probe is that which was observed on the total alder DNA Southern blot (for selected

enzymes). For example a 3.1 kb BamHI fragment hybridized to both alder total DNA and to λ Ano4-1 DNA for the pAnod5 probe. However, in some instances this is not the case. This may be due to the fact that the alder library was constructed from partial MboI digests which may have eliminated certain restriction enzyme sites flanking the hybridizing sequence (cf. the BamHI fragment hybridizing to pAnod4 in λ Ano4-1 is 25 kb which includes the lambda vector left arm (Figure 6) whereas it is approximately 6 kb on the total DNA Southern blots (Figure 4)). Therefore, comparisons of hybridizing fragments of cloned DNA to those of total genomic DNAs can be useful to determine whether there is congruity between selected sequences; however, care must be taken as the fragment sizes detected may not be exactly the same.

The sequences contained within λ Ano4-1 and λ Ano4-2 hybridize to probes derived from both pAnod4 and pAnod5 in distinct regions of the clone (Figure 7) ie. for λ Ano4-1 digested with HindIII (lanes 13) fragments hybridizing to pAnod4 are different from the fragment hybridizing to pAnod5, but all three fragments are from the same cloned alder DNA. The starred fragment in the BamHI/EcoRI lanes (lane 3) appear to be the same size in both autoradiograms. These were found to be two distinct fragments, initially as judged by superimposition of the two autoradiograms, and

later by restriction enzyme site mapping and sequencing. Other independent isolates of λ Anod4 (λ Anod4-4 and λ Anod4-9) showed similar hybridization patterns with both probes (data not shown). The fragments of λ Anod4 clones hybridizing to pAnod4 and pAnod5 were different. This allowed the determination of the distance between the two distinct regions; it was estimated to be 2 kb between the two starred BamHI+EcoRI fragments (Figure 7). Interestingly, the pAnod4 probe did not hybridize to the λ Anod5-5 genomic clone (selected with the pAnod5 probe). The λ Anod5-5 clone, however, may overlap with the λ Anod4-1 clone in the region hybridizing to pAnod5 only. Therefore, it was proposed that the genes hybridizing to the two cDNA insert probes, were linked, with a maximum distance between them of 2 kb.

It was noted that, as pAnod4 and pAnod5 showed the same pattern of expression as seen on Northern blots (T. Gleeson, M.Sc. thesis), then possibly the genes which encode pAnod4 and pAnod5 could be expressed and regulated in the same way.

FIGURE 6: RESTRICTION MAP OF THE THE ALDER CLONES λ Anod4-1 AND λ Anod4-2 AND THEIR POSITIONS WITH RESPECT TO THE TOBACCO CHLOROPLAST MAP

Restriction sites in the two alder clones λ Anod4-1 and λ Anod4-2 were determined by mapping and sequencing. Restriction sites are: Ba, BamHI; Bg, BglII; E, EcoRI; H, HindIII; K, KpnI; M, MboI; Pv, PvuII; S, SalI; X, XhoI. The sizes of the DNA inserts were estimated to be 18.6 kb for λ Anod4-1 and 22.3 kb for λ Anod4-2. Bar is 1 kb. Probes used for screening are indicated.

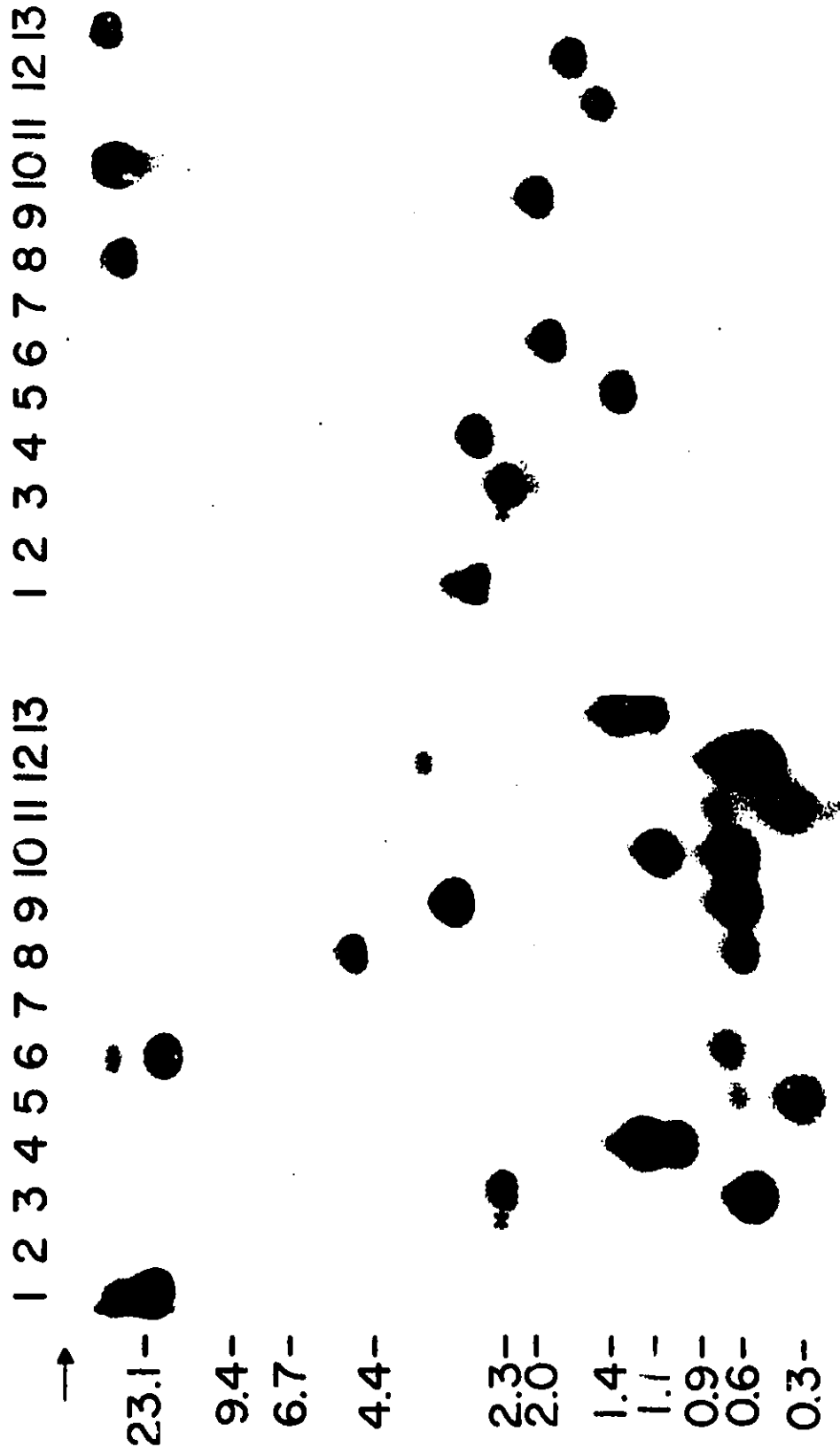
The chloroplast map is derived from Shinozaki et al. 1986b and the IR_A region of the genome with a portion of the SSC region is drawn (see Figure 1). Positions of the cloned DNA on the tobacco chloroplast map were assigned by numbering in the homologous regions of the chloroplast DNA when the sequence analyses were done. JSA denotes the junction point between IR_A and SSC (Shinozaki et al. 1986b).

FIGURE 7: LINKAGE OF TWO ALDER GENES

A Southern transfer of DNA from the alder clone λ Anod4-1 was hybridized sequentially to A: pAnod4 insert probe; B: pAnod5 insert probe. Digestions (1 μ g) were done with 1, BamHI; 2, BamHI+BglIII; 3, BamHI+EcoRI; 4 BamHI+HindIII; 5, BamHI+MspI; 6, BamHI+XhoI; 7, marker lambda DNA + HindIII and ϕ X174 + HaeIII, 8, EcoRI; 9, EcoRI+BglIII; 10, EcoRI+HindIII; 11, EcoRI+MspI; 12, EcoRI+XhoI; 13, HindIII. The sizes (in kb) to the left of the panels represent the molecular weight standard described in lane 7. The arrow indicates the origin. Fragments indicated with an (*) are discussed in the text.

A

B



VI. SUBCLONING AND SEQUENCING

A. SUBCLONING INTO pTZ18R AND M13mp19 VECTORS

1. Strategy of cloning

The smallest hybridizing fragments which would be large enough to map easily and small enough to sequence with relative ease were subcloned. Having available a simple restriction map, fragments outlined in Figure 8 were subcloned into either M13 and/or pTZ primarily for sequencing and further mapping, respectively. Attempts were made to clone each fragment in both orientations. Those clones used in the sequencing reactions are outlined in Figure 8.

2. Choice of vectors

M13mp19 was selected to facilitate the construction of deletions by the Dale procedure (Dale *et al.*, 1985) as the enzyme EcoRI has been found to restrict the DNA in a more reproducible manner. The other alternative would have been to use M13mp18 but this would have required the use of HindIII and it has been documented that cutting is less efficient (IBI: The Cyclone System, version 9.5). M13 phage vectors have been routinely employed for the preparation of single-stranded templates. pTZ18R (Pharmacia) was used as it has the same polylinker as M13mp19 and single-stranded

molecules can be generated from the plasmid by making use of the f_1 origin.

3. Nomenclature of subclones

In order to facilitate data accumulation, clones were named according to the scheme outlined here. Plasmid clones were named with B, E, EB, Bg (representing the digestion from which the fragment originated eg. B, BamHI; E, EcoRI; Bg, BglII), 1, 2, 5 (the isolate number of the lambda phage eg. 1, λ Anod4-1; 2, λ Anod4-2; 5, λ Anod5-5), and Letter (for the specific subclone selected). eg. B2A: would be the first clone (A) selected from the subcloning of a BamHI fragment from λ Anod4-2 into pTZ18R. M13 Phage clones were named in the same manner except that names were preceded by an M (for M13 phage) eg. ME1U: would be the 20th clone (U) selected from the subcloning of an EcoRI fragment from λ Anod4-1 into M13mp19. The only exceptions are N4EB clones (N4A/N4C) and N5EB clone (N5A), which were named in preliminary experiments under another scheme. These were labeled according to N, "nodulin" clones; 4 or 5, the hybridization to their respective probes, pAnod4 or pAnod5; EB denote EcoRI+BamHI fragment. The abridged versions were used to denote orientations of the fragment with A and B representing each orientation. eg. N4A is one orientation of the N4EB clone.

4. Confirmation of clone identity

Plasmid clones were digested with the enzymes used in the cloning, or others designed to cut out the insert in order to verify insert sizes by agarose gel electrophoresis. Hybridization to the proper probes was then done to confirm homology. In the case of M13 phage clones, complementation tests (C-tests) were performed with the phage DNA to determine the relative orientation of the clones and to isolate clones which have the insert in either orientation. Some of the gels were subjected to blotting and subsequent hybridization to verify the clones.

B. SEQUENCING OF CLONED FRAGMENTS

Specific fragments were cloned in both orientations in order to obtain sequence information for both strands. The strategy is outlined in Figure 9, as well as the number of nucleotides determined for each clone. Deletions were prepared to generate overlapping clones in order to facilitate sequencing without subclonings. Phage supernatants were electrophoresed (DIGE) to size inserts, and the gel was blotted to locate the clones that hybridized to the original cDNA probes. Suitable representatives were chosen for sequencing. On average, 40 clones were obtained per 2.5 kb fragment to be sequenced.

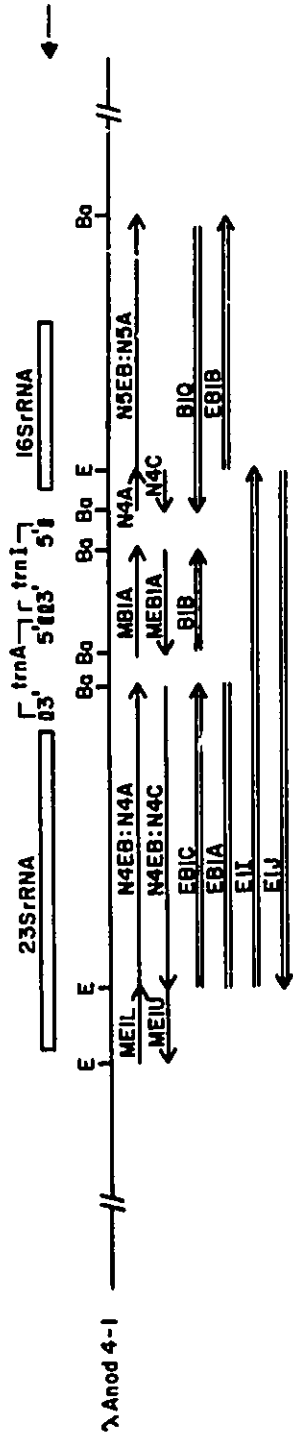
The subclones containing restriction fragments of the

λ Anod4-1 clone outlined in Figure 8 were grouped into three sections to facilitate further analyses with the aid of comparisons of the nucleotide sequence of each clone to the sequence of the tobacco chloroplast genome. The first was named CF423 (Figure 9A): cp, chloroplast sequence; 4, hybridizing to the pAnod4 probe; containing sequences homologous to 23S rDNA. It includes the 0.7 kb EcoRI and the 2.5 kb EcoRI+BamHI fragments that hybridize to pAnod4 insert probe (clones MEIU, ME1L, N4A, N4C). These clones also code for the 3' exon of a tRNA gene, trnA and part of its intron. The second was named CPTRN (Figure 9B): cp, chloroplast sequence; trn, which contains sequences homologous to tRNA genes trnA (5' exon) and trnI (3' exon) and portions of their introns. These sequences hybridized to neither probe used in this work. It includes the 1.0 kb BamHI fragment dividing the two hybridizing regions (clones MB1A and MEB1A). The third region was given the name CP516 (Figure 9C): cp, chloroplast sequence; 5, hybridizing to the pAnod5 probe; 16, which is homologous to 16S rDNA. It includes the 3.0 kb BamHI fragment which hybridizes to pAnod5 insert probe (clones N5A, N4A, N4C). The clones N4A and N4C contained two EcoRI+BamHI fragments from different parts of the map that were cloned and sequenced as one fragment: the complete 2.5 kb EcoRI+BamHI fragment

FIGURE 8: CLONING STRATEGY AND SUBCLONE IDENTIFICATION

The subclones outlined in this figure are named according to the scheme described in the text. Maps presented here are portions taken from Figures 6 and 14 (Appendix 2).

Restriction enzyme sites are: B, BamHI; E, EcoRI; Bg, BglII; X, XhoI. Bar is 1 kb. Double slashes indicate that only a portion of the clones is drawn. Asterisks indicate clones for which there is no sequence data. Arrows indicate the relative orientation of the clones on the original maps. Double lines define plasmid clones, single lines M13 phage clones. Clones N4A, N4C and N5C were used to prepare deletion clones for sequencing.



CP423 CP516

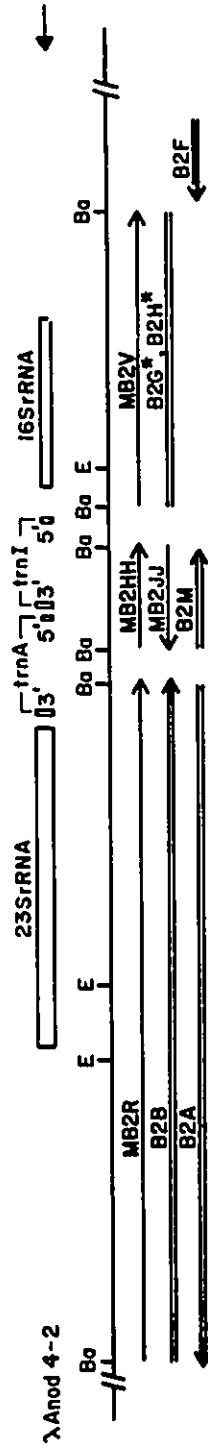
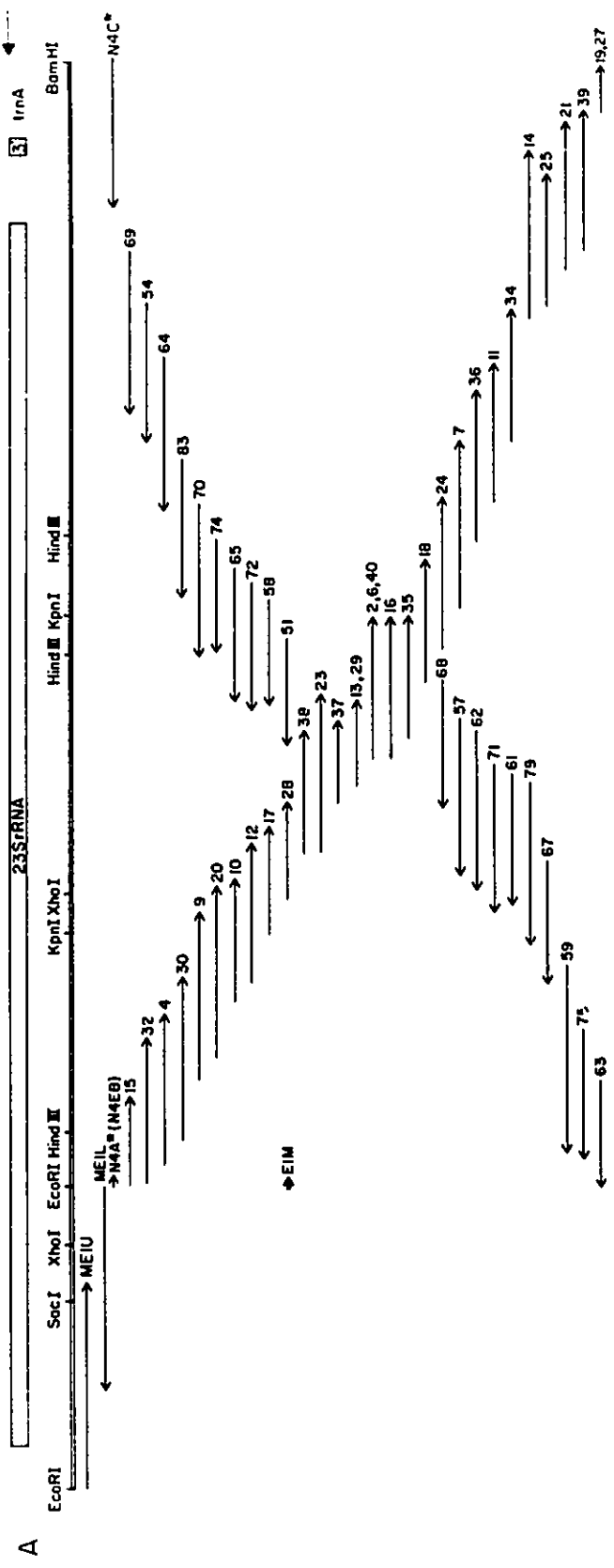


FIGURE 9: SEQUENCING STRATEGY FOR SUBCLONES DERIVED FROM
 λ Anod4-1

Subclones derived from λ Anod4-1 were sequenced according to the methods described in this work. Arrows indicate the orientation of the clones. Restriction enzyme sites are: B, BamHI; E, EcoRI; Bg, BglII; H, HindIII; K, KpnI; S, SacI; X, XhoI. Bar is 0.5 kb. Single lines denote single-stranded templates (M13mp19), double lines, double-stranded templates (pTZ18R). Dotted lines on the maps mark regions for which there is no sequence data. The length of each line corresponds to the number of nucleotides sequenced from that particular clone. Numbers indicate the specific deletion clone sequenced.

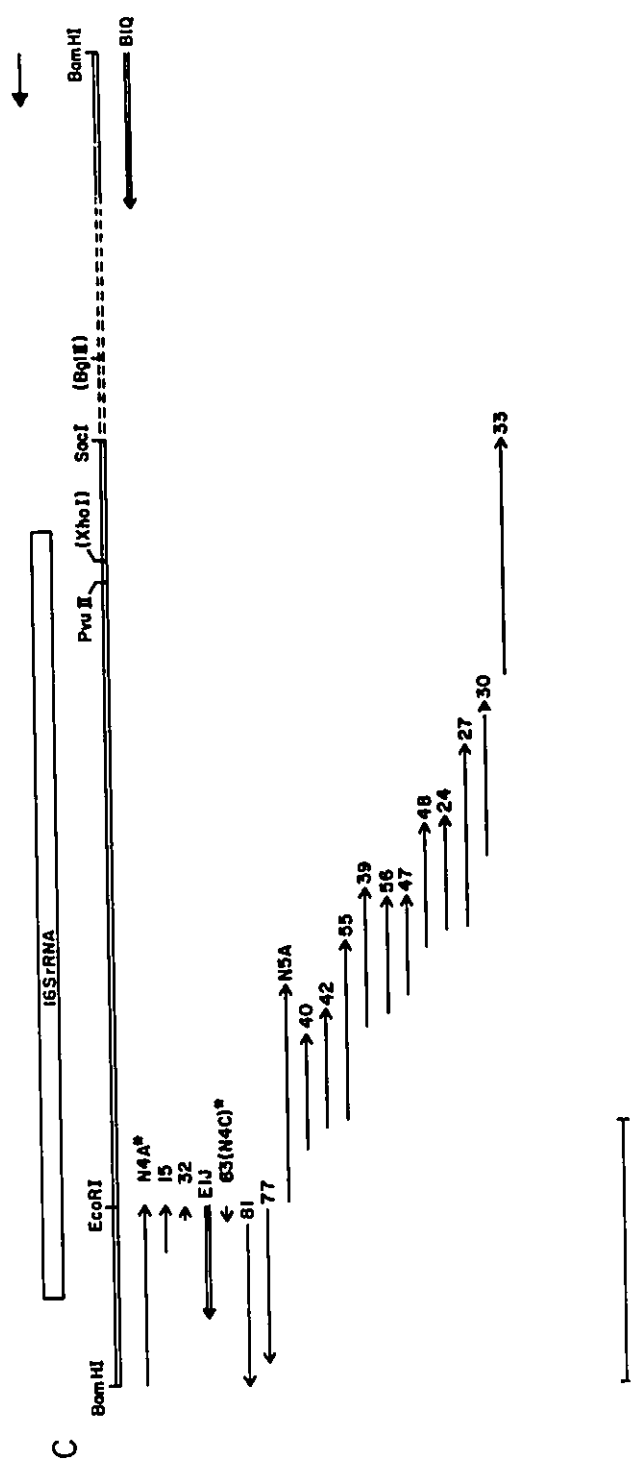
A: CP423 subclones span the length of the 23S rDNA, trnA (3' exon) homologous region. Deletion clones of N4A are oriented from left to right and those of N4C are in the other direction.



—

B: CPTRN subclones partially span the region between the 23S and 16S rDNA genes. The length of the region for which no sequence is available is approximately 40 bp.

C: CP516 subclones encompass the 16S rDNA, trnV, and ORF70B region. The clone N5A was used to generate the deletion clones outlined in this figure in one orientation only. Restriction enzyme sites between brackets were mapped at those locations but no sequence data yet supports their assignment.



hybridizing to pAnod4 and a ~0.4 kb EcoRI+BamHI fragment which mapped within the 3.0 kb BamHI fragment hybridizing to pAnod5.

VII. DNA SEQUENCE ANALYSIS

In summary, an estimated 7.0 kb of the alder clone λ Anod4-1 was subcloned in plasmid and/or M13 phage vectors. Approximately 6.3 kb of nucleotide sequences were obtained which show homology to 16S rDNA, 23S rDNA, and two tRNA genes (trnA and trnI) of the chloroplast genome. The nucleotide sequence obtained from the clones described in Figure 9 are presented in Figure 10. The positions of the probes pAnod4 and pAnod5 used in these studies were also determined by sequence comparisons of the cDNA insert sequences with the tobacco chloroplast genome and illustrated in Figure 6. pAnod4 is found in the 3' portion of the 23S rRNA gene, whereas pAnod5 is located within the 5' end of the 16S rRNA gene.

The length of the sequences actually obtained for CP516, CPTRN, and CP423 are 2112, 847 and 3295 nucleotides, respectively. Upon comparisons with tobacco, coding regions for the alder homologues were assigned. CP516 contains the 16S rDNA (Figure 10A: 954-2438: 1485 bases) and possibly other genes (trnV and ORF70B) but insufficient sequence data

does not allow accurate identification. CPTRN (Figure 10B) codes for part of the trnI intron (1-506: 506 bases), the trnI 3' exon (507-542: 36 bases), the trnA 5' exon (604-641: 38 bases) and part of the trnA intron (642-887: 246 bases). Finally, CP423 (Figure 10C) codes for part of the trnA intron (1-198: 198 bases), the trnA 3' exon (198-233: 36 bases) and the 23S rDNA (391-3201: 2811 bases).

Upon sequence comparisons to plant and organellar data banks, the extent of homology of the alder sequences was the greatest for the tobacco chloroplast sequences. Therefore, the tobacco nucleotide sequence was used to assign the start and stop of coding regions in alder. Of course, these borders could differ in alder by as little as a few nucleotides (as is the case between tobacco and maize) to substantially more.

The nucleotide sequences obtained for the major portion of the clones were from the inverted repeat A (IR_A) (Shinozaki *et al.*, 1986b). In support of this assignment subclones derived from λ Anod4-2 (namely MB2R and B2B) show homology to the ORF1244 in the short single copy region of the tobacco chloroplast genome. It is not known from which repeat clone λ Anod4-1 is derived because nucleotide sequences for both inverted repeats IR_A and IR_B are identical and the nucleotide sequences obtained are well within the

inverted repeat segment. For λ Anod4-1 and λ Anod4-2, homologies to both inverted repeats were detected by computer analysis but only one repeat was diagrammed and used in the analysis. The high degree of homology in sequence of the cloned alder DNA to the chloroplast genome sequence of tobacco (>95%) (and others) must indicate that chloroplast DNA was cloned.

An other alternative to the above explanation is that the homologies of the alder sequences were detected to chloroplast genes inserted into the nuclear genome of alder. Such genetic transfer has been documented, especially in mitochondrial DNA where both nuclear and chloroplast sequence information have been found. Promiscuous chloroplast sequences are widely distributed in plant mitochondrial DNA seemingly in random fashion both with respect to their variation in different plant species and the portion of the chloroplast genome incorporated (Stern and Palmer, 1984). Chloroplast fragments containing two tRNAs and the 16S ribosomal RNA gene (Stern and Lonsdale, 1982) and the ribulose-1,5-bisphosphate carboxylase large subunit are found in maize mitochondria (Lonsdale, 1983). The transfer of these sequences has been thought to have occurred recently since the chloroplast and mitochondrial sequences have a high degree of homology (Levings and Brown, 1989). Usually transferred genes are not functional in

their new location, but some transferred tRNA genes are transcribed, matured and may be used in the mitochondria (Maréchal *et al.*, 1987).

In addition, mitochondrial and chloroplast sequences have been found in nuclear DNA (Newton, 1988; Kemble *et al.*, 1983; Scott and Timmis, 1984; Timmis and Scott, 1983). Therefore it is possible that chloroplast DNA can be found in the nuclear genome of alder since as much as 12 kb (Stern and Lonsdale, 1982) has been reported being transferred. The likelihood of such a transfer is low, and given the amount of chloroplast DNA present in total DNA preparations it is probable that chloroplast DNA from chloroplasts was cloned and analysed. Further analysis of other portions of the lambda genomic clones isolated may define whether or not genes from other compartments are found on the pieces of DNA cloned.

A. NUCLEOTIDE SEQUENCE EVALUATION

The sequence for the portion of CP516 homologous to the 16S gene is almost complete except for two areas (Figure 10A: 964-967; 1273-1286) at the 5' of this gene. The first area was designated because two consecutive deletion clones sequenced in this region did not yet overlap. By sequence comparisons to tobacco, it was estimated that approximately

13 nucleotides were missing. The second area is also defined by homology to the tobacco sequence; only four bases are missing and observation of the sequencing gels shows a slight compression in the gel in this region. A third region 5' to the cloned EcoRI+BamHI fragment is also missing, as described by the dashed lines in Figure 9C, and corresponds to nucleotides 296-782 in Figure 10A. Encoded in this last missing portion would be the trnV gene and the ORF70B gene (by comparison to tobacco). In the case of CPTRN, the sequence is from one strand only (see Figure 9B) but reproducible (repeated three times). By comparison to the restriction map of λ Anod4-1 and to the nucleotide sequence of tobacco rDNA, there are ~40 nucleotides missing in the middle portion of the clones (Figure 10B: 410-451). For both CPTRN and CP516, clones are available to generate future deletions, and obtain the missing nucleotides. The DNA sequence obtained from CP423 is from both strands (Figure 9A) except for three short regions (Figure 10C: 359 to 432, 2590 to 2830, and 3128 to 3295) which are of one strand only. For the first single-stranded region in CP423 overlapping deletion clones in one orientation only were sequenced.

FIGURE 10: NUCLEOTIDE SEQUENCES OF CP516, CPTRN AND CP423
AND SEQUENCE COMPARISONS OF ALDER TO MAIZE AND TOBACCO

The nucleotide sequences for the alder clones were determined using the strategy diagrammed in Figure 9. Restriction enzyme sites used in the cloning and others to serve as reference points are indicated as well as the genes which are encoded by those sequences. Restriction enzyme sites are: A, AccI; B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; Hp, HpaII; K, KpnI; M, MboI; Pv, PvuII; S, SacI; X, XhoI. (<) indicates the direction of transcription and ([or]) the putative start and end of the coding regions.

The complete nucleotide sequence for alder is found in the upper strand, the maize (Edwards and Kössel, 1981; Koch et al., 1981; Schwarz and Kössel, 1981; Schwarz et al., 1981; Strittmatter et al., 1985) and tobacco (Shinozaki et al., 1986b) sequences are presented respectively in the two strands below. The same bases are represented by (.), deletions by (-), and insertions are indicated in lower case letters below the primary sequences with an arrow (^).

A: Nucleotide sequence of CP516. The genes encoded by these sequences include 16S rDNA, trnV and ORF70B. Brackets indicate the number of missing nucleotides, at positions 296, 964, and 1273 as described in the text. Underlined nucleotides indicate putative promoter sequences. No sequence comparisons were available for the first 781 nucleotides of this sequence; therefore, only the tobacco sequence is shown for these positions. The maize sequence is taken from the GenBank file MZECPRG16 and begins at position 456, whereas the tobacco sequence is from the file TOBCPCG and begins at position 101823.

CCCTTCTCAG TGAGTCCTTT CTTTTTTTAT GAACTGTTGG CACCAGTCCT ACATTTTGTC 60
...GA..T..---- ------G.. ...T.....

TCTGTGGACC GAGAAGAAAG GGGGCTCAGC GGAAGAGGA TTGTAATATG AGA²ACGCAA 120
..... ..G.....AG.. G....CC... ..G.A....

GGAGGTCAAC CTCTTTCAAA TATACAACAT ^{Hp}GGATCCGGC AATGCAATGC AATGTAGTTG 180
..... ..T.-- ------..

GACTCTCATG TCGATTCGAA TGAATCATCT ^{Hp}TTCCACCGG AGGTAAATCT TTGCCCTGCT 240
..... ..C....C-... ..-.....

AGGCAAGAGG ATAGCAAGTT ACAAATCTG TCTCGGTAGG ACATGTATTT CTATT(486) 781
..... ..C.....

TGAGAATGGA TAGAGCTCGT ^SGGATTGACGT ⁻³⁵GAGGGGTATG ⁻¹⁰ATGGCTATAT TTCTGGGAGC 841
CGG..... ^ ^ T.. ^ ^ ta - .G. T.....C .G....TG..
..... ^ ^ ^ ^ C.G.
..... a g g

GAACTCCAGG CGAATCTGAA GCGCATGGAT ACAAGTTATG CCTTTGAATG AAAGACAATT 901
..... ^ -G...G.
.....G..A.....G.....

CCGAATCCGC TTTGTTTACG AACAAGGAAG CTATAAGTAA TGCAACTATG AATCTCATGG 961
.....C.... ..T..... ..
.....C.... ..

16S rRNA

1022
 AGA(4)GATC CTGGCTCAGG ATGAACGCTG GCGGCATGCT TAACACATGC AAGTCGGACG
 ... ^ A...
 gttc
 ... ^
 gttc

1082
 GGAAGTGGTG TTTCCAGTGG CGGACGGGTG AGTAACGCGT AAGAACCTGC CCTTGGGAGG
A.....

PV 1142
 GGAACAACAG CTGGAAACGG CTGCTAATAC CCCGTAGGCT GAGGAGCAAA AGGAGGAATC
A T..... ..A.....

1202
 CGCCTGAGGA GGGGCTCGCG TCTGATTAGC TAGTTGGTGA GGCAATAGCT TACCAAGGCG
CA.....
C.....

M M 1262
 ATGATCAGTA GCTGGTCCGA GAGGATGATC AGCCACACTG GGACTGAGAC ACGGCCCAGA

1331
 CCTCTACGGG A(13)AATTT CCGCAATGGG CGAAAGCTGA CCGAGCAATC GCGGTGAGGT
 .TC..... ^t ^c ^gcTG.....
 .TC..... ^t ^gc ^c ^g

Hp 1391
 AGAAGGCCAT AGGCGTGCTA CTTCCTTTCC CGGAGAAGAA GCAAGTACGG TATCTGGGAA
 G.....TA C..-.CG.. ^A.....T A...TG.... ^a.....
CA C..-.CG.. ^A..... ^gTG.... ^g

Hp/1451

TAAGCATCGG CTAACTCTGT GCCAGCAGCC GCGGTAATAC AGAGGATGCA AGCGTTATCC
.....G.....
.....

1511

GGAATGATTG GGCCTAAAGC GTCTGTAGGT GGCTTTTTAA GTCCGCCGTC AAATCCCAGG
.....C.....
.....

1571

GCTCAACCCT GGACAGGCGG TGGAAACTAC CAAGCTGGAG TACGGTAGGG GCAGAGGGAA
.....
.....

1631

Hp M

TTTCCGGTGG AGCGGTGAAA TCCGTAGAGA TCGGAAAGAA CACCAACGGC GAAAGCACTC
.....A.T.....
.....

1691

TGCTGGGCCG AACTGACAC TGAGAGACGA AAGCTAGGGG AGCGAATGGG ATTAGATACC
.....A.....G.....
.....

1751

CCAGTAGTCC TAGCCGTAAA CGATGGATAC TAGGCGCTGT GCGTATCGAC CCGTGCAGTG
.....T.....AC.....
.....

1811

CTGTACGTAA CGCGTTAAGT ATCCCGCCTG GGGAGTACGT TCGCAAGAAT GAAACTCAA
.....GC.....
.....GC.....

1871

GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAT GCAAAGCGAA
.....G.....
.....

1931

GAACCTTACC AGGGCTTGAC ATGCCGCGAA TCCTCTTGAA AGAGAGGGGT GCCTTCGGGA
.....
.....

1991
ACGCGGACAC AGGTGGTGCA TGGCTGTCGT CAGCTCGTGC CGTAAGGTGT TGGGTTAAGT
.....
.....

2051
CCCGCAACGA GCGCAACCCT CGTGTTTAGT TGCCACCGTT GAGTTTGAA CCCTGAGCAG
.T..... .TA.-A...
.....T....A...

2111
Hp Hp
ACTGCCGGTG ATAAGCCGGA GGAAGGTGAG GATGACGTCA AGTCATCATG CCCCTTATGC
.C..... T.....A... .G.C..
.....

2171
Hp M
CCTGGGGCAG ACACGTGCTA CAATGGCCGG GACAAAGGGT CGCGATCCCG CGAGGGTGAG
.....G... ..T..
.....-

Hp/2231
CTAACCCCAA AAACCCGTCC TCAGTTCGGA TTGTAGGCTG CAACTCGCCT GCATGAAGCC
.....T....C.....A
.....C.....

2291
Hp E Hp
GGAATCGCTA GTAATCGCCG GTCACCGATA CGGCCGGTGAA TTCGTTCCCG GGCCTTGATC
.....G.C... ..C... .C.....
.....G.C... ..C.....

2351
ACACCGCCCG TCACACTATG GGAGCTGGCC ATGCCCGAAG TCGTTACCTT AACCACAAGG
.....AG.TTT.... ..A...
.....G.....

2411
AGGGGGATGC CGAAGGCAGG GCTAGTGAAGT GGAGTGAAGT CGTAACAAGG TAGCCGTAAGT
.....T.....TA.T.C....
.....

_____]

M

2471

GGAAGGTGCG GCTGGATCAC CTCCTTTTCA GGGAGAGCTA ACGCTTGGTTG GGTATTTTGG
T....A...
 ^tcgta
T.....

2531

TTTGACACTG CTTACACCCC AAAAAGAAGC GAGCTACGTC TGGGTAAAC TTGGAGATGG
G... ..G C..... ..A.C.... ..T....
 ^cc ^aaa
A.....

2591

AAGTCTTCTT TCGTTTCTCG ACGGTGAAGT AAGACTAAGC TCATGAGCTT ATTTTCCTAG
A-- --..... ..C.... ..A.....
C..... ..C..... ..A.....

B 2614

GTCGGAACAA GTTGATAGGA TCC
 A..AG.T-. .AG
 C..

B: Nucleotide sequence of CPTRN. This sequence codes for the trnA intron and 5'exon of trnA, as well as the 3' exon of trnI and a portion of its intron. These areas are defined on the sequence. Within the brackets at position 410 are indicated the number of missing nucleotides. The maize sequence is taken from the GenBank file MZECPRG16 and begins at position 2684, whereas the tobacco sequence is from the file TOBCPCG and begins at position 104800.

B Hp Hp X 60
 CGATCCGGGC GGTCCGGGAG GGACCACCAT GGCTCCTCTC TTCTCGAGAA TCCATACATC
 ^.....G. ..G.....GC
 ^.....G.C
 ^.....G.C

X 120
 CCTTATCAGT ATATGGACAG CTATCTCTCG AGCACTGGTT AGGTTTCGGC CTCAATGGGA
 G.....G..A..... ^.....-..T.
 G.....A..... ..----- --.....

trnI intron

M 180
 AAATAAAACG GAGCACCTAA CAACGTATCT TCACAGACCA AGAACTACGA GATCGCCCCT
-....C....
T.C....A.....

M Hp240
 TTCATTCTGG GGTGACGGAG GGATCGTACC ATTCGAGCCT TTTTTTTCAT GCTTTTCCCG

GT. ^AA.GGA.
 ^a ^cga ^g

300
 GAGGTCTGGA GAAAGCTGCA ATCAATAGGA TTTCCTAAT CCTCCCTTTC CGAAGGAAGA
 .C.....A.T.C..C.....C. T. ^a.....
 C....T..A. A..G..A.CTT .GAGT ^gtcG ..GGG.C.GG AGGGT..C.T AACGCC----

360
 AGTTGAAATT CTTTTTCCTT TCAGCAGGGA CCAGGAGATT GGACTAGCCA TAAGAAGAAT
 ^c.....T..... ^t.....G....
 -----..T... ^c...T...AG TT.TTTC.CA ^a.....T..... ^ggg ^g.....G

457
 GCTTGGCTGA TAATAACTCA TTCTTGGCTT CGACCCTCAG TCACTACGA(40) GTCGAGT
-..-.. ^a.....GC.. ^c..... ^t.....C..... ^ta..... A.T..-.
 .GG-..AAC. AGCAC...TG GAGAGC..AG TACAA.GG.. AGTTGTATG A.T..-.

_____] [> _____
 M/Bg 517
 ACTGCTCCAA AAGATCTATA TCTCTCCATG TGCTAGTCAT ATTACTTACG GCAAGGTCTC
 T...-.....C- ..-T.GGT.G GA...-GT.G. G...-GA.G.. -..... ^c ^g .G.....
 T...-.....C- ..-T.GGT.G G.C.-GT.G. ...-GA.G.. -..... ^c ^g .G.....

_____ trnI 3' exon _____]
 TGGTTCAAGT CCAGGATGGC CCCCATGCGC CAGTAAAAGA ATAGAAGAAG CATCTGACTC 577
 ^t ^g ^g .G.....A
 ^tgc ^g .G.....A

[> _____ trnA 5' exon _____
 S 637
 CTTTCATGCAT GCTCCACTTG GCTCGGGGGG ATATAGCTCA GTTGGTAGAG CTCGGCTCTT
 ^t A..... ^g

_____] [> _____ trnA intron _____
 697
 GCAATGGGT CGTTGCGATT ACGGGTTGGA TGTCTAATTA TCCAGGCGAT AATGATAGTA
CGG.
GG.

_____ Hp 757
 TCTTGTACCT GAACCGGTGG CTCACITTTT CTAAGTAATG GGAAAGAGGA CCGAAACATG
G..... .T.....
G..... .G.....

_____ 817
 CCACTGAAAG ACTCTACTGA GACAAAGATG GGCTGTCAAG AACGTAGAGG AGGTAGGATG
C ^aaA ..G.....

_____ M/Bg M 877
 GGCAGTTGGT CAGATCTAGT ATGGATCGTA CATGGACGGT AGTTGGAGTC GGCGGCTCTC
A.....

B 887
 CTAGGGATCC
CT...
 .C....T...

C: Nucleotide sequence of CP423. The genes encoded by this sequence include a portion of the chloroplast trnA intron, the 3' exon of trnA, as well as the chloroplast 23S rRNA. These areas are defined on the sequence. This section is 3295 nucleotides in length. The maize sequence is taken from the GenBank files MZECPRG16 between positions 4156-4533 and MZECPRRG beginning at position 1, whereas the tobacco sequence is from the file TOBCPCG and begins at position 105938.

960
TGAAATAGAA CATGAAACCG TAAGCTCCCA AGCAGTGGGA GGAGCCTGGG GCTCTGACCG
.....G.....G.....G.AAA.T. A.....
.....ctg.....t.....A.....

1020
CGTGCCTGTT GAAGAATGAG CCGGCGACTC ATAGGCAGTG GCTTGGTTAA GGAACCCAC
.....tgga

1080
Hp
CGGAGCCGTA GCGAAAGCGA GTCTTCATAG GGCAATTGTC ACTGCTTATG GACCCGAACC
.....G.....

1140
M H
TGGGTGATCT ATCCATGACC AGGATGAAGC TTGGGTGAAA CTAAGTGGAG GTCCGAACCG
.....A.....CA.....

1200
ACTGATGTTG AAGAATCAGC GGATGAGTTG TGGTTAGGGG TGAAATGCCA CTCGAACCCA

1260
GAGCTAGCTG GTTCTCCCCG AAATGCGTTG AGGCGCAGCA GTTGA CTGACTGGA CATCTAGGGG

1320
K
TAAAGCACTG TTTCGGTGCG GGCCGCGAGA GCGGTACCAA ATCGAGGCAA ACTCTGAATA
.....T...C.....

H1380
CTAGATATGA CCTCAAATA ACAGGGGTCA AGGTCCGCCA GTGAGACGAT GGGGGATAAG
.....CA.....T.....

1440

HpM

CTTCATCGTC GAGAGGGAAA CAGCCCGGAT CACCAGCTAA GGCCCCTAAA TGACCGCTCA
.....T..
.....T.....

1500

GTGATAAAGG AGGTAGGGGT GCAGAGACAG CCAGGAGGTT TGCCTAGAAG CAGCCACCCT
.....G.....A.....T.....
.....

1560

M

TGAAAGAGTG CGTAATAGCT CACTGATCGA GCGCTCTTGC GCCGAAGATG AACGGGGCTA
.T.....CT.-.....C.....T.....
.....

1620

M

AGCGATCTGC CGAAGCTGTG GGATGTAAAA AGGCATCGGT AGGGGAGCGT TCCGCCTTAG
.....C...T.....
.....TA.....

1680

A

GGGAAGCAC CTGCGCGAGC AGCGGTAGAC GAAGCGGAAG CGAGAATGTC GGCTTGAGTA
A.....A AC...AA...G.G...C...
A.A.....CT .C.....G.T...G...

1740

ACGCAAACAT TGGTGAGAAT CCAATGCCCC GAAAACCTAA GGGTTCCTCC GCAAGGTTCC
...A.....C..T.....
.....

1800

M

TCCACGGAGG GTGAGTCAGG GCCTAAGATC AGGCCGAAAG GCGTAGTCGA TGGACAACAG
.....TA.....
.....

1860

GTGAATATTC CTGTACTGCC CCTTGTTGGT CCCGAGGGAC GGAGGAGGCT AGGTTAGCCG
..C.....A...A.G.....
.....

1920

AAAGATGGTT ATCGGTTCAA GGACGCAAGG TGCCCCTGCT TTTCAGGGT AAGAAGGGGT
.....A...T...A.....A.....
.....A...T...T.....G.....

X 1980
AGAGAAAATA CCTCGAGCCA ATGTTTCGAGT ACCAGGCGCT ACGGCGCTGA AGTAACCCAT
.....GG .G..C..... ..A..... G.A...GA..
tgagccccgtggactagccattgcttctccacgaggctcataccaggcgtacggcggaagtat
gta
.....GA.

K 2040
GCCATACTCC CAGGAAAAGC TCGAACGACC TTCAACAAAA GGTACCTGT ACCCGAAACC
.....
.....T ..G.G...G.

2100
GACACAGGTA GGTAGGTAGA GAATACCTAG GGGCGCGAGA CAACTCTCTC TAAGGAACTC
.....G
.....G

2160
GGCAAAATAG CCCCCTAACT TCGGGAGAAG GGGTGCCTCC TCACAAAGGA GGTGCGAGTG
.....C.. ..G...G
.....G

Hp 2220
ACCAGGCCCG GCGACTGTT TACCAAAAAC ACAGGTCTCC GCAAAGTCGT AAGACCATGT
.....A

Hp 2280
ATGGGGGCTG ACGCCTGCCC AGTGCCGGAA GGTC AAGGAA GTTGGTGACC TGATGACAGG
.....A.

Hp 2340
GGAGCCAGCG ACCGAAGCCC CGGTGAACGG CGGCCGTAAC TATAACGGTC CTAAGGTAGC
.A...G...
.....G...

M 2400
GAAATTCCTT GTCGGGTAAG TTCCGACCCG CACGAAAGGC GTAACGATCT GGGCACTGTC
.....
.....

2460

TCGGAGAGAG ACTCGGTGAA ATAGACATGT CTGTGAAGAT GCGGACTACC TGCACCTGGA
 G.....
 G.....

2520

H

CAGAAAGACC CTATGAAGCT TCACTGTTCC CTGGGATTGG CTTTGGGCCT TTCCTGCGCA
T.....
T.....

2580

Hp

GCTTAGGTGG AGGGCGAAGA AGGCCTCCTT CCGGGGGGGC CCGAGCCGTC AGTGAGATAC
A.....
A.....

2640

E

CACTCTGGAA GAGCTAGAAT TCTAACCTTG TGTCAGGACC TACGGGCCAA GGGACAGTCT
G.....
G.....
 ^
 C

2700

A

CAGGTAGACA GTTTCTATGG GGCCTAGGCC TCCCAAAGG TAACGGAGGC GTGCAAAGGT

2760

X

TTCCTCGGGC CAGACGGAGA TTGGCCCTCG AGTGCAAAGG CAGAAGGGAG CTTGACTGCA
C.....
G.....

2820

M

AGACCCACCC GTCGAGCAGG GACGAAAGTC GGCCTTAGTG ATCCGACGGT GCCGAGTGGA
T.....
A.....

2880

M

AGGGCCGTCG CTCAACGGAT AAAAGTTACT CTAGGGATAA CAGGCTGATC TTCCCCAAGA

2940

S

GCTCACATCG ACGGGAAGGT TTGGCACCTC GATGTCCGGCT CTTGCCACC TGGGGCTGTA
 .TC.....
A.....

In all but one case, compression of DNA sequence observed using the dGTP termination mixes was removed successfully when dITP mixes were substituted in the sequencing reactions.

The nucleotide sequence of each region was evaluated for the base composition of the actual sequence obtained, ie. the missing sections were not included in the analysis. The percent G+C content of each of the genes was calculated from the nucleotide sequence, for the sequences obtained as is presented below. The coding regions in alder genes for these RNA species were determined from sequence comparisons with tobacco. Table 3 summarizes the values of G+C content for the RNA genes and sequenced flanking regions in alder.

In general, the 23S and 16S rDNA sequences are G+C rich compared to the whole chloroplast genome which is said to be A+T rich (Wolfe and Sharp, 1988). In liverwort, it was reported that there was a 52.6% G+C content for ribosomal RNA coding sequences and 52.1% G+C in the coding sequence for tRNAs (Ohyama et al., 1988b; Ozeki et al., 1989), as opposed to 28.8% G+C overall. Tobacco chloroplast overall G+C content is slightly higher than that of liverwort (Ozeki et al., 1989). The G+C content for alder sequences seem to show this trend as well, as the coding regions for the ribosomal RNA genes and the transfer RNA

TABLE 3: G+C content of the nucleotide sequence obtained from the alder clone λ Anod4-1

<u>Alder gene</u>	<u>Length of sequence</u> (bases)	<u>G+C content (%)</u>
<u>Coding</u>		
23S rRNA	3295	55.0
16S rRNA	1467	56.2
trnA (5'/3'exons)	38/35	56.2
trnI (3' exon)	36	61.1
<u>Non-coding</u>		
5' of 16S rRNA + 3' of 16S rRNA	643	43.8
3' of trnI and 5' of trnA	61	49.2
trnI intron	467	46.9
trnA intron	444	49.0
5' of 23S rRNA and 3' of trnA + 3' of 23S rRNA	250	48.6

genes are above the 50 percent G+C as opposed to the portions which separate the coding regions which fall below the 50 percent mark. In contrast, mitochondrial spacer DNA is not A+T rich but has approximately the same base composition as the bulk of the mitochondrial DNA which is ~47% G+C (Gray, 1989b).

Again by comparison to tobacco sequences, it was determined that two small BamHI fragments between CP423 and CPTRN, and CPRN and CP516 were initially overlooked in the restriction analysis, and were not subcloned. Upon re-analysis of digestions performed with E1J subclone, it was found that there was more than one small fragment generated by cleavage with EcoRI+BamHI. These fragments are indicated in Figure 6. These two fragments encode portions of the trnA and trnI introns and the 5' trnI exon, as determined by sequence comparisons with tobacco.

Some subclones derived from λ Anod4-2 were also partially sequenced. These subclones are shown in the map in Figure 6. Preliminary sequencing identified these as identical to subclones from λ Anod4-1 and therefore sequencing was not pursued further. Clones B2A, MB2HH, MB2JJ, MB2V (Figure 8) were identical to clones from λ Anod4-1. The clone MB2V shows homology to the N5A clones but the homology stops after about 191 nucleotides. The reason for this break in homology is not known. Both clones were derived from fragments of the same size isolated independently from both lambda clones; perhaps the sequence of other clones spanning this area may yield useful information. Comparisons of MB2V to tobacco generated no other similar sequences in the genome, ruling out recombination at repeated sequences. Clones B2B and MB2R

also derived from λ Anod4-2 share homology in sequence but have no corresponding sequences in λ Anod4-1 (see Figure 6). The B2F subclone shows homology to tobacco chloroplast sequences upstream of the 16S rDNA at another putative BamHI site. This analysis shows that independent clones selected with the same probe could have the same genomic structure without obvious rearrangements.

B. PRIMARY NUCLEOTIDE SEQUENCE ANALYSIS: COMPARISONS TO OTHER ORGANISMS

The order of the ribosomal genes is very well conserved throughout kingdoms. In chloroplasts, the ribosomal RNA genes are found in the order 16S-trnI-trnA-23S-4.5S-5S (5'>3') in both inverted repeat regions; therefore, there are two copies per genome. The gene order in tobacco chloroplast DNA is found in most other angiosperms, in at least one fern and one gymnosperm (Strauss *et al.*, 1988), and differs by only a single inversion from the liverwort order (Ozeki *et al.*, 1989). Sequence comparisons were performed against a number of genes within plant, organelle and bacterial data banks. For the sake of simplicity, only those homologies significant with this work were retained. It must be noted that although the rice chloroplast genome has been sequenced (Hiratsuka *et al.*, 1989), comparisons could not be performed as the complete nucleotide sequence

was not available for use in the data bank at Agriculture Canada. Because the nucleotide sequence homologies were the highest to those of tobacco chloroplast, assignment of the gene borders of the alder sequence were done by comparison to the tobacco sequences.

1. 16S rDNA

Sequence comparisons of the alder clones were performed against plant, bacterial and organelle data banks (alignments not shown). The percent homology of the alder 16S rRNA gene to other genes are as follows: *E. coli*, 75.1% (Brosius *et al.*, 1978), *A. nidulans*, 83.4% (Tomioka and Sugiura, 1983), liverwort, 94.2% (Kohchi *et al.*, 1988), maize chloroplast, 95.0% (Schwarz and Kössel, 1980), and tobacco chloroplast, 97.9% (Tohdoh and Sugiura, 1982).

The homologies for the coding and flanking regions were also calculated for the sequenced portions of the 16S rDNA of alder with those of maize and tobacco. The aligned sequences of these three genes and flanking regions are presented in Figure 10A. The 16S rDNA alder coding sequence (1485 bp) was 95.0% and 97.9% homologous to the maize (1490 bp) and tobacco (1489 bp) sequences, respectively. The % value obtained did not include the missing nucleotides outlined in Figure 10C. The homologies to the flanking

regions of both published chloroplast genes were also determined. The 5' flanking region (nucleotides 1-952 of CP516, not including missing nucleotides in the count), is 93.6% and 91.0% homologous to the maize (172 bp) and tobacco (468 bp) 5' regions, respectively. In addition, the alder 16S rRNA gene 3' flanking region of 175 bp (nucleotides 2439-2614 of CP516) is 87.6% and 93.3% homologous to maize (177 bp) and tobacco (180), respectively. As anticipated, the percent homology of alder to tobacco sequences is greater in the coding region (Schaal and Learn, 1988) rather than in the non-coding region.

From comparisons to chloroplast rrn promoter consensus sequences derived from maize, tobacco, spinach, duckweed and mustard (Strittmatter *et al.*, 1985), putative -35 and -10 sequences were defined on the alder sequence. The -10 region is defined by nucleotides 827-832 and the -35 region by 805-810 of CP516 (Figure 10A). The consensus sequences for rrn promoters are summarized in Table 4. The extent of homology of the alder 16S "promoter sequences" to the chloroplast consensus sequence is 83% (38 matches out of 46). In this respect, the alder putative promoter shows a greater degree of homology to the chloroplast rrn consensus sequence than that of *A. nidulans* or *E. coli*, which is what was expected. At the nucleotide level the -10 sequence is much more conserved between the members which are included

in Table 4, than are the -35 sequences. DNA sequences underlined in the chloroplast promoter consensus sequence represent essential promoter elements *ctp1* (-35 sequence) and *cpt2* (-10 sequence) that affect transcription when deleted or mutagenized (reviewed in Gruissem *et al.*, 1989b). Similar sequences have been found in alder and it remains to be determined whether these will be efficient for transcription initiation. Promoter strengths can be

TABLE 4: COMPARISON OF *rrn* PROMOTER CONSENSUS SEQUENCES

	-35	-10
cp:	GTGGGA <u>TTGACG</u> TGA--GGGtAGGG-TgGC <u>TATAtT</u> -	CTGGgaGGg
an:	AGGGGG GTGACG CGACTAGGC GAGTTAGG TAGATTA ATTAAGCGC	
ec:	tt-c-c TTGt-- -----acTCCC	TATAATG CGCC-CCa-
ai:	GTGG-A TTGACG TGA-GGGGTATGA TGGC TATATTT	CTGGGAGCG

Homologous nucleotides in all samples are indicated by capital letters and those in four chloroplast samples out of five and five *E. coli* samples out of six by small letters. Dashes indicate nucleotides vary at this position. Table taken from Kumano *et al.*, 1986. *cp*, chloroplast *rrn* conserved sequences; *an*, *A. nidulans* *rrnA* sequences, *ec*, *E. coli* P1 conserved sequences; *ai*, *Alnus incana*.

determined with reporter genes and chloroplast extracts and chloroplast run-on transcription systems, as has been done for *psbA*, *trnM2*, *rbcL* and *atpB* genes (reviewed in Gruissem, 1989b). The relative strengths are presented from greater to lesser, in the order of the promoters listed.

From other work, it was found that the 3' terminal region of the liverwort 16S rRNA gene also contains the sequence CCTCCT, which is complementary to prokaryotic SD sequence (Shine and Dalgarno, 1974). The existence of a sequence complementary to the SD sequence in 16S rRNA is a feature of prokaryotes, and this may influence the translation efficiency of mRNA (Kohchi *et al.*, 1988). This same sequence is found in alder and tobacco sequences for 16S rRNA at the 3' end of the coding region (positions 2431-2436 on CP516: Figure 10A) by analogy to the liverwort sequence. Ribosomal RNAs are involved directly in mRNA binding by a specific interaction of a sequence at the 3' end of the 16S ribosomal RNA with a complementary sequence on the chloroplast mRNAs which is located a few nucleotides upstream of the AUG codon (Steinmetz and Weil, 1989).

2. *trnI*, *trnA*

Analyses of sequence homologies for alder *trnI* and *trnA* tRNA genes have been determined with respect to tobacco and maize and are presented in Figure 10B and 10C. The 5' exon

of trnA (Figure 10B: n 604-641) is 100% homologous to that of maize and tobacco (both 38 bp in length). It is deduced from this comparison that the alder trnA 5' exon will also be of the same length. The 3' exon of the alder trnA (Figure 10C: n 199-233) is 97.3% homologous to its maize counterpart (37 bp) and 100% homologous to the tobacco gene (35 bp). The maize exon starts and ends one nucleotide 5' and 3' relative to the tobacco exon. Therefore, the borders of the trnA 3' exon indicated on the figure for the alder gene have been assigned by comparison to tobacco, since alder chloroplast sequences have an overall greater homology to the tobacco sequences.

No sequence comparison of the trnI 5' exon was produced as this exon is found within a small fragment that was not cloned (see Figure 8: the ~400 bp BamHI fragment). The 3' exon of trnI in alder (Figure 10B: n 507-542) was 86.5% homologous to the maize exon (37 bp), but 88.6% homologous to the tobacco exon (35 bp) in a pairwise comparison. Again, the maize trnI 3' exon borders are defined outside the tobacco borders at positions 506-544 on the alder sequence (Figure 10B).

The nucleotide sequence obtained in the region of bases 240-575 (Figure 10B) would require further confirmation by

using overlapping clones for sequencing, as the sequence information is in one strand only and does not yet overlap.

Intron homologies could not be carried out completely with tobacco because portions were within two small BamHI fragments that were not subcloned. The portions of the trnA intron that were sequenced (444 bp: n 642-887 of CPTRN and 1-198 of CP423), however, show 92.9% and 93.0% homology to the maize and tobacco sequences, respectively. The portions of the alder trnI intron that were sequenced (467 bp: n 1-410 and 451-506 of CPTRN) are 98.7% and 65.5% homologous to the maize and tobacco sequences, respectively. The extent of homology to maize in this instance is noteworthy: its significance is not known since the homology of the other alder sequences to the tobacco sequences have been greater than that to maize. In addition, the region separating the 5' exon of trnA and the 3' exon of trnI (Figure 10B: n 543-603) has been sequenced (61 bp) and it shows 85.2% and 88.5% homology to the maize (63 bp) and tobacco (64 bp) sequences. Therefore, the trend is maintained; homologies between coding regions are greater than in non-coding regions, except for the short spacer region separating trnA 5'exon and trnI 3'exon. Since none of the tRNA genes encode the mature CCA sequence at its 3' terminus (Kohchi *et al.*, 1988), it is anticipated that this will also be the case for alder.

For both tRNA genes, intron sizes cannot be calculated accurately but from comparison to maize and tobacco, intron sizes should range between 708 n (tobacco) and 949 n (maize) for the trnA intron and between 708 n (tobacco) and 806 n (maize) for the trnI intron. Unfortunately, only about 47-63% of the sequence of the trn gene introns has been determined, depending on the plant chloroplast trn gene used for comparison. As group II introns (Michel and Dujon, 1983) in tRNA genes are characteristic of the chloroplast genome of land plants (Ohyama et al., 1988), it is expected that the type of introns detected in alder will be the same since the primary nucleotide sequences shows varying degrees of homologies to 70 published group II-like introns (Michel et al., 1989). Analysis of one small region provides additional support. The nucleotide sequences obtained from alder were analysed to search for any obvious repeated elements. One short sequence of 47 nucleotides within CP423 was repeated, that is there was homology detected to two separate regions in both the maize and tobacco sequences, at the 3' end of the trnA and trnI introns. The 47 nucleotides from 84-130 of CP423 (Figure 10C) are part of the trnA intron, and are identical to both maize and tobacco sequences in this same intron. These same sequences are found repeated in the trnI intron upstream from the start of the 3' exon of trnI in maize and tobacco. Unfortunately, in

alder the second portion of the repeat within the trnI intron, ends at position 455 in the CPTRN (Figure 10B) sequence which borders those nucleotides that are missing. When compared to tobacco chloroplast sequences, the 47 nucleotide sequence from alder trnA intron shows 81.3% homology to the region within the trnI intron. This specific 47 nucleotide sequence forms a portion of domain IV, all of domain V, and part of domain VI in the conserved secondary structure of Group II-like introns (Michel *et al.*, 1989). This domain constitutes a hairpin loop which is the best conserved structural element of group II introns, playing a central role in their assembly (Jarrell *et al.*, 1988). The primary sequence and secondary structure of domain V are sufficiently well conserved and distinctive to be used as determinants of group II introns (Michel *et al.*, 1989). Therefore, it is likely that the alder sequences serve the same purposes in the processing of introns in the two tRNA genes.

trnA and trnI are found in the 23S-16S rDNA spacer region of many bacteria including *A. nidulans* (Kumano *et al.*, 1986).

3. 23S rDNA

The homologies of the cloned alder sequences to those of other organisms have been calculated using paired aligned

comparisons (alignments not shown). Homologies of the 23S rRNA gene of alder with the following organisms were noted: *E. coli*, 66.5% (Brosius *et al.*, 1980), *Anacystis nidulans*, 70.1% (Kumano *et al.*, 1983), liverwort chloroplast, 94.1% (Kohchi *et al.*, 1988), maize chloroplast, 95.3% (Edwards and Kössel, 1981) and tobacco chloroplast, 98.3% (Takaiwa and Sugiura, 1982), not counting the 3'-terminal regions of the *E. coli* and *A. nidulans* 23S rRNAs which code for the 4.5S found in land plant chloroplasts. Sequence alignments of the alder clone containing the 23S rRNA gene and flanking regions, with those of the maize and tobacco chloroplasts are found in Figure 10C.

The coding region of the 23S rDNA of alder (2811 bp; Figure 10C) has been calculated to be 95.3% and 98.3% homologous to maize (2881 bp) and tobacco 23S rDNA (2810 bp), respectively. As expected, the regions flanking this gene in alder show a lesser extent of homology to the maize and tobacco sequences. The 157 nucleotides in the 5' flanking region (nucleotides 234-390 of CP423) are 77.6% and 93.6% homologous to maize (156 n) and tobacco (153 n), respectively. The 94 nucleotides at the 3' end of the 23S rDNA (nucleotides 3202-3295 of CP423) are 83% and 94.7% homologous to the maize (100 n) and tobacco (98 n) sequences, respectively. The start of the maize structural 23S rRNA is two nucleotides 5' to that of tobacco and ends 6

nucleotides 5' to that of tobacco. In summary, the sequences are very well conserved between alder and tobacco in the coding region, but not as conserved in the regions flanking this gene.

The primary structure of the 23S rRNA gene in alder and tobacco is quite simple, as opposed to that of C. reinhardtii for which the coding region is interrupted by a group I intron (Allet and Rochaix, 1979). This intron encodes an ORF which is structurally related to mitochondrial reading frames that potentially code for maturases (Rochaix et al., 1985). Others report intron sequences in chloroplast 23S ribosomal genes such as that of Chlorella (Yamada and Shimaji, 1987).

The extent of homology of the alder sequence to the chloroplast sequences for both 23S and 16S rRNA genes as well as two tRNA genes demonstrates that these sequences are most likely chloroplast in origin. These genes are cotranscribed and extensively processed to form functional molecules, perhaps such conservation is required for the proper recognition of enzymes responsible for the processing.

C. CONSERVED SECONDARY STRUCTURE

As previously stated, the regions noted as CP516, CPTRN and CP423 encode the genes for, or parts of the 16S rRNA, trnI, trnA and 23S rRNA in alder chloroplast.

Like tRNAs, the 23S (or large subunit) and 16S (or small subunit) rRNAs have elaborate secondary structures over the length of the molecule. The overall structure is separated into domains as described earlier. Within these domains are conserved nucleotide sequences which form the universal core of the molecule. Most of the published data on secondary structures is based on the *E. coli* 16S rRNA model proposed by Woese (Woese *et al.*, 1989; Gutell *et al.*, 1986; Gutell *et al.*, 1985; Woese *et al.*, 1983). Universal regions are common to all organisms as opposed to variable regions which range in length and structure between organisms. The loops and helices depicted in this secondary structure interact with the 30S ribosomal proteins to form the small subunit of the ribosome.

Secondary structure models can be generated by the simple pencil and paper method. Many structures have been reported in the literature, and primary sequence alignments of two sequences, one which constitutes the reference sequence on the published model and the other the "unknown"

FIGURE 11: SECONDARY STRUCTURE OF THE ALDER 16S rRNA

The alder 16S rRNA secondary structure was derived from the superimposition of the changes in sequence of alder observed with respect to maize. Sequence alignments (Figure 10A) were used as a basis for the construction of the alder structure. The maize chloroplast sequence is taken from Gutell *et al.*, 1985 p. 164. Open squares denote deletion of base(s), (>) indicate insertions, filled arrows indicate one base change in alder leading to compensatory base pairing; open arrows indicate two base changed in alder which are compensated. Boxed nucleotides represent those nucleotides for which the sequence was not yet determined (Figure 10A). The asterisk denotes an ambiguity observed in the maize sequence: sequence from data banks indicate a C residue at this position whereas the model has incorporated a U.

sequence, allow for enough similarities for the secondary structure to be drawn. E. coli 16S and 23S rRNA models (Figure 3) are currently available for use but these sequences are more distantly related to the chloroplast sequences. Gutell and associates (Gutell et al., 1985) have published the secondary structure of the maize chloroplast 16S rRNA which is more easily comparable to the alder gene than is the E. coli structure, given the primary alignment of the sequences (Figure 10A). Figure 11 illustrates the secondary structure of the maize chloroplast 16S rRNA onto which are indicated the alder sequences. Bases that are the same in alder and maize were left unchanged, bases that are different (either additions, deletions or changes) are indicated.

Overall, the alder and maize secondary structures are identical, with a few modifications. In ten instances (Figure 11, filled arrows), the change in one base in alder with respect to the maize sequence allows for better Watson-Crick base pairing as opposed to the non-compensating pairs which were present before. Likewise, twenty changes in bases result in a non Watson-Crick base pair. In addition, compensatory base pairing in both strands occurs in two cases (Figure 11, open arrows). At these positions, the base on one strand forming the helix and the "pairing" base on the other strand in this same helix are both changed

allowing for classical C/G, A/U pairing. These constitute what is referred to as compensatory base changes (Woese et al., 1983). The other base changes mostly occur in the single stranded portions of the molecule or are the result of inversions of bases at particular positions. A more detailed comparison of part of this structure with those of other secondary structures will be presented in the Discussion section. It must be noted that the nucleotide sequence obtained for the 16S rRNA gene in alder is mostly single-stranded through overlapping clones. Therefore, the degree of certainty of the sequence is lesser than that which it could be if both strands had been sequenced, although nucleotide sequences were verified many times from the sequencing gels.

This preliminary analysis of the alder secondary structure of the 16S rRNA molecule is in agreement with the concept that even if two primary sequences are divergent for the rRNA molecules, their secondary structures will retain a conserved core with variable regions surrounding it. Further analysis could be performed with the chloroplast 23S rRNA molecule but this would lengthen the present text substantially. It is anticipated, however, that the overall structure would be very similar to that observed for maize, with compensating and non-compensating base changes throughout the molecule.

CHAPTER 4: DISCUSSION

In summary, an alder genomic library has been constructed and 16S and 23S ribosomal RNA genes as well as portions of two tRNA genes and their introns have been isolated. Sequence analysis has revealed that the λ Anod4-1,-2 clones are most homologous to the chloroplast 16S and 23S ribosomal DNA and tRNA genes as opposed to the λ Anod5-1 clone which shows homology to the mitochondrial 18S ribosomal DNA of maize, soybean and wheat (λ Anod5-5 is discussed in Appendix 2).

I. ALDER SEQUENCES AND PHYLOGENY

Chloroplast DNA has been described by Palmer (Palmer *et al.*, 1988a) as an "extremely valuable molecule for phylogenetic studies". To date, the growing ribosomal RNA database now includes sequences from organisms representing all three primary lines of descent; eubacteria, archaebacteria and eukaryotes. In addition, because the rate of fixation of the mutations is low, the "conservatively evolving ribosomal RNA genes hold the greatest promise for resolving the deepest branching of plant evolution and have already settled the ultimate question of chloroplast evolution, namely, its endosymbiotic origin" (Palmer *et al.*, 1988a). For studies of phylogenetic

levels higher than the family level ie. order, analysis of the distribution of major structural rearrangements, such as inversions and the loss or gain of genes and introns can be sufficient to extract phylogenetic information from chloroplast DNA (Palmer et al., 1988a). Other studies with chloroplast RFLPs can extend phylogenetic trees to the species level, but obtaining ribosomal DNA sequences is thought to be more appropriate for study of the genus and species taxonomic levels even considering the cost of sequencing. Ribosomal RNAs are the most useful data set from which to make inferences about organellar origins because they are ubiquitous among the various genomes which are to be analysed (Gray, 1989a).

Without resorting to the sophisticated preparation of phylogenetic trees, it is more than likely that alder chloroplast sequences branch with the tobacco chloroplast genome on the unrooted phylogenetic tree inferred from a small or large subunit rRNA sequence data base (Figure 2). The deduced structure of the alder ribosomal RNA operon in addition to the extent of homology of the alder sequences to the plant chloroplast sequences (~98% homology in the coding regions of the 16S and 23S rDNA to tobacco) would support its proposed position in phylogeny. On the primary branch from which maize, tobacco, C. reinhardtii chloroplasts separate, Anacystis, a cyanobacterium, is found (Figure

2A). This result suggests that chloroplasts are most closely related to cyanobacteria, supporting the endosymbiotic theory. Also, in chloroplasts, not only gene sequence but also gene organization and mode of expression provide strong evidence for chloroplasts originating from cyanobacteria (Gray, 1989a).

The secondary structures for the 16S rRNA for E. coli, A. nidulans, maize/alder mitochondria and chloroplasts are strikingly similar, despite their obvious differences at the primary nucleotide level (data not shown). The structure spanning positions 378-706 of the E. coli model were selected for further analysis since the alder sequences from both mitochondria (18S mitochondrial rRNA gene: Figure 15) and chloroplast (16S chloroplast rRNA gene: Figure 10A) were available from this work.

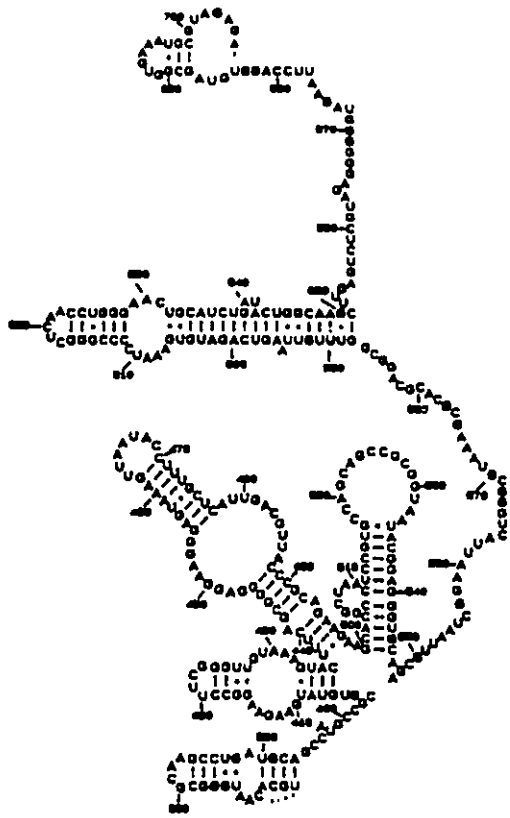
Upon comparison of the four structures in Figure 12, it is possible to define both conserved and variable regions which constitute part of the small subunit rRNA domains II and III.

Three conserved, universal regions located from positions 515-536, 569-575, and 688-699 (Gutell et al., 1985) are part of the core structure ie. the structure most highly constrained phylogenetically which is presumably most critical for assembly and function of small subunit

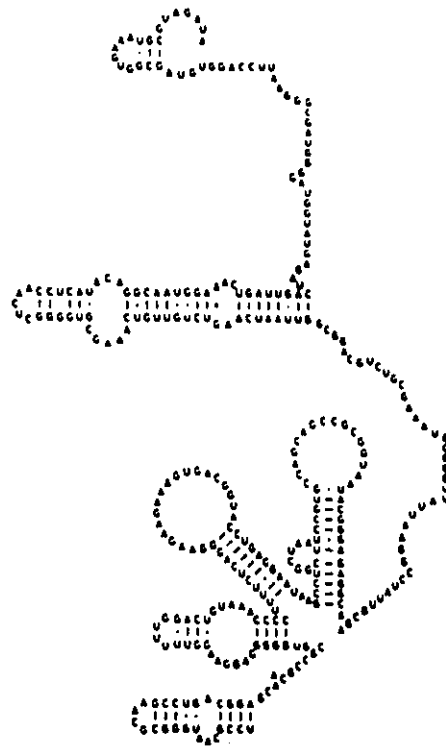
FIGURE 12: COMPARISON OF THE SECONDARY STRUCTURES OF 16S rRNA MOLECULES AMONG BACTERIA, CYANOBACTERIA, MITOCHONDRIA AND CHLOROPLASTS.

The secondary structures for the 16S rRNA molecule of E. coli (a), Anacystis nidulans (b), maize mitochondria (c) and maize chloroplast (d) are presented. The alder nucleotide sequence obtained for the mitochondrial homologue (MBg50 subclone; Figure 15) and chloroplast homologue (n 1301-1603, in CP516; Figure 10A) are included on the maize mitochondrial (c) and chloroplast (d) figures. Notations used are those outlined in Figure 11. (a) is copied from Noller et al., 1987 p. 697; (b), (c) and (d), from Gutell et al., 1985 pp. 163, 167 and 164, respectively.

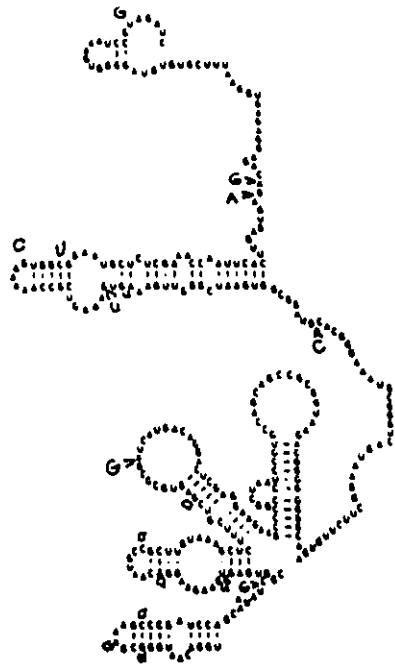
a



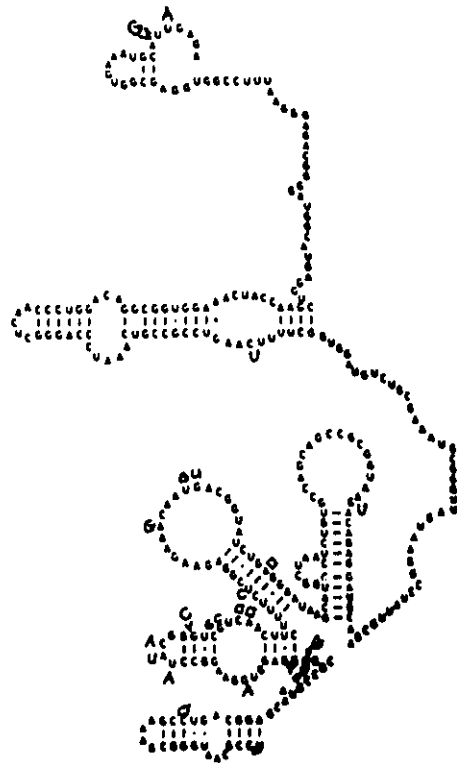
b



c



d



ribosomes (Gutell et al., 1985). Positions 530 and 693, among others, are suggested sites for tRNA and/or mRNA contacts; whereas positions 674, 703, and 705 are potential 50S subunit contacts (Gutell et al., 1985). The cytosine residue at position 507 is extremely well-conserved and is potentially involved in tertiary base-base interactions with the guanine residue at position 9 (Gutell et al., 1985; Gutell et al., 1986; Gutell and Woese, 1990). The structure of residues 500-545 is that of a compound helix containing a side loop. Eubacteria seem to always contain a bulge of six bases starting at residue 505. The entire structure (500-545) seems extremely important to ribosome function as it is found in all organisms, organelles and the sequence is highly conserved, almost universal in the apex loop (Woese et al., 1983). Part of its importance stems from its role in ribosomal protein binding: the S4 protein binds near position 500 in the upper part of the 5' domain, in a region around which are clustered five helices. S4 protein is considered a primary binding protein, along with others that are required for the binding of subsequent proteins (Noller et al., 1987). Almost all stem and loop structures in the secondary structure for maize (alder) chloroplast 16S RNA are observed in the A. nidulans 16S rRNA sequence through the aid of compensating base changes (Tomioaka and Sugiura, 1983).

In addition to the conserved regions, variable portions of the molecule are also noted. Some 40% of the molecule is not completely conserved in its secondary structure (Gutell et al., 1985); some of these sections may however be conserved within phylogenetic groups (defining kingdom-specific motifs) or they may show variation even within phylogenetic groups (Gutell et al., 1985). The number of nucleotides between positions 404-499 varies, but eubacteria, chloroplasts, archaebacteria and some mitochondria have two compound helices in this region, though different in form in each of these groups (Gutell et al., 1985). In mitochondria and chloroplasts as well as in A. nidulans, the helix found in positions 455-477 is deleted entirely with respect to the E. coli structure (Woese et al., 1983). Positions (588-651) form the protein S8 binding loop and helix. Again, the apex loop is highly conserved (Woese et al., 1983). The S8 protein is also a primary binding protein and it makes widespread contacts and/or has a major effect on the confirmation of the central domain of the rRNA structure (Noller et al., 1987). In addition, the helix is progressively shortened in the mitochondria starting with the plants, where a single base is deleted and finishing with mammalian mitochondria where the entire helix is deleted. It was reported that the mitochondrial rRNA structure appears to represent a "stripped down version of the translation apparatus" (Woese et al., 1983). This is

certainly the case for some mitochondria; however, mitochondrial ribosomal RNA are very heterogenous in size (Raué *et al.*, 1988). Differences in size between eubacterial and chloroplast 16S-rRNAs can be partially accounted for by changes of the number of nucleotides in the 436-500 structure (Raué *et al.*, 1988).

Thus similarities of the secondary structures of the 16S rRNAs provide additional information as to the phylogenetic relationships between bacteria, mitochondria and chloroplasts.

II. USE OF CLONED SEQUENCES AS PROBES TO DIFFERENTIATE BETWEEN CELLULAR DNA FROM DIFFERENT COMPARTMENTS

The work described in this thesis may help to define specific probes to distinguish between the four genomes involved in symbiosis namely: bacterial, plant nuclear, plant mitochondrial and plant chloroplast. Two types of probes could be defined: one type to differentiate between the genomes of each compartment, another to differentiate between kingdoms eg. plant or bacterial. Heterologous probes require sufficient conservation of base sequence to ensure a significant level of cross-hybridization and linear arrangement of the chromosome to permit alignment of the cross-hybridizing fragment and interpretation of the hybridization signals (Palmer *et al.*, 1988a).

Presently, ribosomal DNA sequences from each compartment have been detected by sequence homology to genes from data banks. The pAnod4 (23S rDNA) probe used in this work shows a high degree of similarity (as determined by the IBI programs) to bacterial ribosomal DNA in general, but the degree of match to for example, E. coli and Bacillus subtilis is 71.4% and 74.6%, respectively; therefore, it was proposed that the ribosomal DNA originated from Frankia, the actinomycete which forms symbiotic associations with alder. Sequence analysis for pAnod4 is more significant in this respect because of the length of the insert (1718 bp) is greater than for pAnod5. Sequence homologies of pAnod5 also point to bacterial sequences (data not shown).

The other probe used at the start of this work was demonstrated to have homology to plant nuclear 25/26S ribosomal DNAs. Data bank searches with the pAnod2 insert sequences showed a significant degree of homology to Citrus limon 26S, tomato 25S, and rice 25S ribosomal genes (using the Wisconsin package of the NRC). In the course of the present work, cross-hybridization of the pAnod4 insert probe to the pAnod2 insert DNA was observed in control dots. Therefore, under the conditions used, cross-hybridization of the sequences is possible.

It was determined by sequence homology to data banks that the BglIII subclone of λ Anod5-5 was most likely plant mitochondrial 18S ribosomal DNA. The highest scores were obtained for soybean, maize, and wheat 18S rRNA genes, as described in Appendix 2. The alder clone may be a representative of this cellular compartment's genome. It is important to note that unlike chloroplast ribosomal genes, only the 18S and 5S rRNA genes are tightly linked in mitochondrial genomes and the 26S rRNA gene lies elsewhere in the mitochondrial genome (Levings and Brown, 1989).

Finally, λ Anod4-1,-2 clones and subclones showed the highest homology to either plant chloroplast 23S or 16S ribosomal RNA genes. The extent of homology to tobacco is approximately 98% as described in the results section of this work. Thus, it is likely that portions of these clones would be useful in the detection of the chloroplast genome.

In the above description, one point becomes obvious: bacterial sequences hybridized to mitochondrial, chloroplast and nuclear ribosomal RNA genes under usual hybridization conditions (Materials and Methods, II.A.1). To be useful, sequences from the aforementioned clones must be selected and proper hybridization conditions determined to define a specificity for the probes.

The identification of an area which is sufficiently diverged between genomes will be necessary. Recently, it was suggested that rapidly evolving regions of the large subunit mitochondrial ribosomal RNA gene may be useful targets for selection of oligonucleotide probes specific to fungal genes (Bruns and Taylor, 1989). Perhaps such a strategy would be viable in this instance, also. The variable regions in the sequence may be a starting point to suggest sequences to be used. Oligonucleotides have been used extensively to differentiate between sequences (Wallace and Miyada, 1987). As hybridization probes they are very specific as they can detect a single base mismatch when under 23 nucleotides in length (Siebert *et al.*, 1989). Such probes may be selected on the basis of size and number of matches with the ribosomal RNAs in the different genomes. For example, an oligonucleotide of a defined length selected in the chloroplast clones may show a high percent homology to other chloroplast sequences, but low percent homology to nuclear or mitochondrial or bacterial sequences. Similar oligonucleotides could be selected in this way for the genomes of each compartment. It may be necessary to use mixed or multiple oligonucleotide probes in order to allow for sequence divergence of ribosomal genes between cultivars.

Once defined, the utility of this region (oligonucleotide) should be investigated by computer analysis, to first determine the specificity before hybridization analyses are attempted. If the oligonucleotides show relatedness to only one genome, then hybridization studies could proceed. DNA could be isolated from alder mitochondria, chloroplast, nuclei, and Frankia to verify specificity of the probes under different hybridization conditions.

Candidates suggested for specific probes follow. For plant mitochondrial 18S rRNA, nucleotides 183-202 (55% G+C) of the λ Anod5-5 subclone MBg50 (Figure 15) could be useful as preliminary computer sequence analysis shows this 20-mer is 100% homologous to mitochondrial sequences of soybean, maize and wheat only, but has 13, 8, 13 mismatches with liverwort, Anacystis and tobacco chloroplast sequences, respectively (data not shown). Similarly, nucleotides 1479-1498 (55% G+C) of the CP516 sequence are conserved in liverwort and tobacco chloroplast sequences but there are 8 mismatches out of 20 for the Anacystis sequence. Interestingly, the selected sequences would map between two universal regions of a defined universal core (Grabau, 1985) identified for 16S-like ribosomal RNA sequences (Woese et al., 1983). This suggests that these probes may be useful to distinguish between chloroplast and mitochondria in

vascular plants but not Anacystis, within 16S rDNA-like sequences.

Similarly, two regions in the sequence CP423 seem to be chloroplast specific for the 23S rDNA gene in plants. Nucleotides 1057-1076 and 2652-2671 (both 55% G+C) show 100% homology to the 23S rDNA chloroplast genes of tobacco, liverwort and maize but show 6 and 10 mismatches for Anacystis in preliminary computer analyses. A priori, these oligonucleotides may be useful to distinguish between plant and bacterial sequences. At present, a computer analysis is required to determine if these 23S rDNA sequences are useful.

An alternative which would require a major sequencing effort would be to isolate ribosomal rRNA genes from each genome (mitochondrial, chloroplast, nuclear, bacterial), sequence them and align the sequences for comparisons. A simpler way to do this, given proper computer software, would be to align a number of genes already present in data banks along with the sequences in this work for one compartment, determine conserved regions between them and use the sequences as probes.

The probes derived from such a study could be used to search in other genomes. Preliminary results have

demonstrated that pAnod4/5 probes have been used successfully to detect homologous sequences in birch. At low stringency, the bacterial sequences cross-hybridize to the ribosomal RNA of chloroplast, nucleus and mitochondria. It has not yet been determined in which compartment the birch ribosomal RNA genes are located, but the specific probes designed in this work could be used to distinguish between the compartments, if preliminary results are favourable.

From an academic point of view, the isolation and characterization of ribosomal RNA genes selected by the mentioned putative probes could be useful to elucidate phylogenetic relationships between plant species. As previously stated, the sequences for a number of rRNA genes including approximately seven for the SSU and six for the LSU of chloroplasts, have been reported. These alder sequences can add to this data bank and it is the first tree species reported. Increases in the size of the data sets analysed provide a stronger basis for establishing phylogenies. That chloroplasts are derived from cyanobacteria is already established from a variety of data but as new rRNA genes are included in the analyses, perhaps new insights will be gained as to a single or multiple endosymbiotic events. From a practical point of view, these same probes could be used in combination (ie. the probe used

would be a mixture of oligonucleotides from the different compartment) to detect the presence of contaminating ribosomal genes in cDNA libraries prior to the screening for protein coding genes. Alternatively, full-length rDNA cloned inserts could be used for this same purpose. In this last case, the reduction of lost time would be tremendous.

Still another option would be to use highly conserved regions of the ribosomal RNA genes as primer target sites (Lane et al., 1985) for PCR (polymerase chain reaction) and direct sequence determination of ribosomal RNA genes, using the amplified gene as template. This approach has been used for partial nucleotide determination of the Mycobacterium kansasii 16S rRNA gene, using 5' and 3' 20-mer primers flanking this ribosomal gene for PCR and a set of 12 primers for direct sequencing (Edwards et al., 1989). The oligonucleotides were selected following a comparison of 20 prokaryotic, eubacterial 16S rRNA sequences including representatives of the major eubacterial groups (purple photosynthetic bacteria, cyanobacteria, for example).

III. FUTURE WORK

A. Short term

Initially, work should be focused on obtaining the

missing information in the regions CP423, CPTRN and CP516C through sequencing. Most clones are readily available or easily prepared for this exercise and comparisons of the primary sequence with chloroplast gene sequence should confirm the presence of i) the 5' exon of trnI and ii) trnA intron in the two small BamHI fragments that remain to be cloned; iii) portions of the trnI intron within the small number of nucleotides missing in the CPTRN sequence; iv) the presence of trnV and ORF50B in the 486 bases missing in the CP516 sequence, in addition to the few nucleotides remaining to be sequenced in the coding region of the 16S chloroplast rRNA alder sequence.

B. Long term

In order to use the data obtained to generate a phylogenetic position for alder, the sequences for both 23S and 16S rDNA must be made accessible to those laboratories with facilities to do the analyses. Possibly a report in Nucleic Acids Research would suffice to present the nucleotide sequence. The 23S rDNA sequence of alder is complete and could be presented as such, but the 16S rDNA would require the sequencing of additional DNA.

It would be of interest to generate the secondary structure of the 23S rRNA species in alder. Most of the

studies on secondary structure are done on the 16S rRNA although structures for 23S rRNA have been presented (reported in Cedergren et al., 1988). Obtaining the secondary structure of a chloroplast gene with the nucleotide positions on the structure would greatly simplify the task as differences in sequence between alder and *E. coli* (for which the secondary structure map is readily available) are numerous. More sophisticated computer programs would be required to obtain a secondary structure for the 23S and 16S rRNA from alder which can then be compared to those which have been published by Woese (Woese et al., 1983; reviewed in Gutell et al., 1985) and others. By comparison of the alder 16S rRNA secondary structure to that of maize chloroplast 16S rRNA and *E. coli* 16S rRNA it can be demonstrated once more that there is, in all organisms, a conserved core structure with some areas of variability.

From analogy to chloroplast genes and more specifically to tobacco, the genomic clones λ Anod4-1 and λ Anod4-2 contain other rRNA genes (4.5S and 5S) and tRNA genes (trnV, trnN, trnR, trnL), several open reading frames, as well as some protein coding genes such as the genes for the 30S ribosomal proteins 7 and 12 (exons 2 and 3) and NADH dehydrogenase subunit ND2. These genes may also be models for gene evolution studies. Furthermore, some of these genes contain

introns: thus intron splicing mechanisms for trnI and trnA, as well as protein coding genes such as rps12 and ndh5 could be investigated if conserved.

Therefore, a wealth of information can be gathered and inferred from a nucleotide sequence, especially that of rRNA genes as these now play an important role in describing phylogenies as well as describing their functional roles.

Finally as a portion of the alder library was stored amplified, it is possible to screen it for nodulin gene homologues in alder using heterologous probes, which is what was originally proposed for this project.

REFERENCES

- Aguettaz, P., Seyer, P., Pesey, H., Lescure, A.-M. (1987). Relations between the plastid gene dosage and the levels of 16S rRNA and rbcL gene transcripts during amyloplast to chloroplast change in mixotrophic spinach cell suspensions. *Plant Mol. Biol.* 8: 169-177.
- Allet, B., Rochaix, J.D. (1979). Structural analysis at the ends of the intervening DNA sequences in the chloroplast 23S ribosomal genes of *C. reinhardtii*. *Cell* 18: 55-60.
- Baumgartner, B.J., Rapp, J.C., Mullet, J.E. (1989). Plastid transcription activity and DNA copy number increase early in barley chloroplast development. *Plant Physiol.* 89: 1011-1018.
- Belanger, F.C., Hepburn, A.G. (1990). The evolution of CpmpG methylation in plants. *J. Mol. Evol.* 30: 26-35.
- Bendich, A.J. (1987). Why do chloroplasts and mitochondria contain so many copies of their genome? *Bioessays* 6: 279-282.
- Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant DNA. *Nucl. Acids Res.* 7: 1513-1523.
- Bowman, C.M. Dyer, T.A. (1979). 4.5-S ribonucleic acid, a novel ribosome component in the chloroplasts of flowering plants. *Biochem. J.* 183: 605-613.
- Briat, J.F., Dron, M., Mache, R. (1983). Is transcription of higher plant chloroplast ribosomal operons regulated by premature termination? *FEBS Lett.* 163: 1-5.
- Brosius, J., Palmer M.L., Kennedy, P.J., Noller, H.F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 75: 4801-4805.
- Brosius, J., Dull, T.J., Noller, H.F. (1980). Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 77: 201-204.
- Bruns, T.D., Taylor, J.W. (1989). PCR analysis of fungal mitochondrial ribosomal RNA genes. Meeting on the polymerase chain reaction: Methodology and applications, April 3-7, 1989. *J. Cell. Biochem. Suppl.* 13 PART E: 276.

- Cedergren, R., Gray, M.W., Abel, Y., Sankoff, D. (1988). The evolutionary relationships among known life forms. *J. Mol. Evol.* 28: 98-112.
- Chao, S., Sederoff, R., Levings III, C.S. (1984). Nucleotide sequence and evolution of the 18S ribosomal RNA gene in maize mitochondria. *Nucl. Acids Res.* 12: 6629-6644.
- Cutting, J.A., Schulman, H.M. (1969). The site of heme synthesis in soybean root nodules. *Biochem. Biophys. Acta* 192: 486-493.
- Dale, R.M.K., McClure, B.A., Houchins, J.P. (1985). A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* 13: 31-40.
- Dawson, J.O. (1983). Dinitrogen fixation in forest ecosystems. *Can. J. Microbiol.* 29: 979-992.
- Dawson, J.O. (1986). Actinorhizal plants: their use in forestry and agriculture. *Outlook on Agriculture* 15: 202-208.
- Dean, C., Pichersky, E., Dunsmuir, P. (1989). Structure, evolution, and regulation of RbcS genes in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40: 415-439.
- Delaney, T.P., Cattolico, R.A. (1989). Chloroplast ribosomal DNA organization in the chromophytic alga Olisthodiscus luteus. *Curr. Genet.* 15: 221-229.
- Deng, X.W., Tonkyn, J.C., Peter, G.F., Thornber, J.P., Gruissem, W. (1989). Post-transcriptional control of plastid mRNA accumulation during adaptation of chloroplasts to different light quality environments. *The Plant Cell* 1: 645-654.
- Denhardt, D.T. (1966). A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Comm.* 23: 641-646.
- Dormann-Przybyl, D., Strittmatter, G., Kössel, H. (1986). The region distal to the rRNA operon from chloroplasts of maize contains genes coding for tRNA^{Acg}(ACG) and tRNA^{Glu}(GUU). *Plant Mol. Biol.* 7: 419-431.

- Edwards, K., Kössel, H. (1981). The rRNA operon from Zea mays chloroplasts: nucleotide sequence of 23S rDNA and its homologies with E. coli 23S rDNA. Nucl. Acids Res. 9: 2853-2869.
- Edwards, U., Rogall, T., Blöcker, H., Emde, M., Böttger, E. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucl. Acids Res. 17: 7843-7853.
- Elmerich, C. (1984). Molecular biology and ecology of diazotrophs associated with non-leguminous plants. Bio/Techn. 2: 967-978.
- Feinberg, A.P., Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Addendum Anal. Biochem. 137: 266-267.
- Feinberg, A.P., Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.
- Frischauf, A.M., Lehrach, H., Poutstka, S., Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170: 827-842.
- Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H., Ohyama, K. (1988). Structure and organization of Marchantia polymorpha chloroplast genome: III Gene organization of the Large Single Copy Region from rbcL to trnI (CAU). J. Mol. Biol. 203: 333-351.
- Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J., Pace, N.R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. 170: 3584-3592.
- Govers, F., Nap, J.-P., Moermann, M., Franssen, H.J., van Kammen, A., Bisseling, T. (1987). cDNA cloning and developmental expression of pea nodulin genes. Plant Mol. Biol. 8 425-435.
- Grabau, E.A. (1985). Nucleotide sequence of the soybean mitochondrial 18S rRNA gene: evidence for a slow rate of divergence in the plant mitochondrial genome. Plant Mol. Biol. 5: 119-124.

- Graf, L., Kössel, H., Stutz, E. (1980). Sequencing of 16S-23S spacer in a ribosomal RNA operon of Euclena gracilis chloroplast DNA reveals two tRNA genes. *Nature* 286: 908-910.
- Gray, J.C., Phillips, A.L., Smith, A.G. (1984). in *Chloroplast Biogenesis*, R.J. Ellis ed., Cambridge Press, Cambridge, pp. 137-163.
- Gray, M. W. (1989). The evolutionary origins of organelles. *TIG* 5: 294-299.
- Gray, M.W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* 5: 25-50.
- Gray, M.W., Cedergren, R., Abel, Y., Sankoff, D. (1989). On the evolutionary origin of the plant mitochondrion and its genome. *Proc. Natl. Acad. Sci, USA* 86: 2267-2271.
- Gray, M.W., Sankoff, D., Cedergren, R.J. (1984). On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucl. Acids Res.* 12: 5837-5852.
- Gray, P.W., Hallick, R.B. (1978). Physical mapping of the Euclena gracilis chloroplast DNA and ribosomal RNA gene region. *Biochem.* 17: 284-289.
- Greenberg, B.M., Gruissem, W., Hallick, R.B. (1984). Accurate processing and pseudouridylation of the chloroplast transfer RNA in a chloroplast transcription system. *Plant Mol. Biol.* 3: 97-109.
- Gruissem, W. (1989a). Chloroplast Gene Expression: How Plants Turn Their Plastids On. *Cell* 56: 161-170.
- Gruissem, W. (1989b). Chloroplast RNA: transcription and processing. *The Biochemistry of Plants* 15: 151-191.
- Gruissem, W., Deng, X.w., Tonkyn, J. (1989c). Control of chloroplast gene expression during plant development and environmental changes. in *Photosynthesis*, Alan R. Liss, Inc., pp 269-284.
- Gruissem, W., Barkan, A., Deng, X.w., Stern D. (1988). Transcriptional and post-transcriptional control of plastid mRNA levels in higher plants. *TIG* 4: 258-263.

- Gruissem, W., Elsner-Menze., C., Latshaw, S., Narita, J.O., Schaffer, M.A. (1986). A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription. *Nucl. Acids Res.* 14: 7541-7556.
- Gutell, R.R., Woese, C.R. (1990). Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of noncanonical pairs. *Proc. Natl. Acad. Sci. USA* 87: 663-667.
- Gutell, R.R., Noller, H.F., Woese, C.R. (1986). Higher order structure in ribosomal RNA. *EMBO J.* 5: 1111-1113.
- Gutell, R.R., Weiser, B., Woese, C.R., Noller, H.F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. *Progress in Nucleic Acid Research and Molecular Biology*, Academic Press, Inc. 32: 155-216.
- Hanahan, D. (1985) Techniques for Transformation of *E. coli*, Chapter 6 in *DNA Cloning Volume 1: a practical approach*, D.M. Glover ed., IRL Press, Oxford, Washington DC. pp. 109-135.
- Hattori, J., Gottlob-McHugh, S.G., Johnson, D.A. (1987). The isolation of high-molecular weight DNA from plants. *Anal. Biochem.* 165: 70-74.
- Hemmingsen, S.M., Woolford C. van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W., Ellis, R.J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 333: 330-333.
- Hermann, R.G., Possingham, J.V. (1980). Plastid DNA-the plastome. In *Results and Problems in Cell Differentiation: Chloroplasts*, J. Reinert ed., 10: 45-94. Berlin/New York: Springer-Verlag. 240 pp.
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C., Meng, B., Li, Y., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217: 185-194.

- Hori, H., Osawa, S. (1987). Origin and evolution of organisms as deduced from 5S sequences. *Mol. Biol. Evol.* 4: 445-472.
- Hori, H., Lim, B.-L., Osawa, S. (1985). Evolution of green plants as deduced from 5S rRNA sequences. *Proc. Natl. Acad. Sci. USA* 82: 820-823.
- Jarrell, K.A., Dietrich, R.C., Perlman, P. (1988). Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.* 8: 2361-2366.
- Jenni, B., Stutz, E. (1979). Analysis of Euglena gracilis chloroplast DNA: mapping of DNA sequences complementary to 16S RNA outside of the three rRNA gene sets. *FEBS Lett.* 102: 95-99.
- Jorgensen, R.A., Cuellar, R.E., Thompson, W.F., Kavanagh, T.A. (1987). Structure and variation in ribosomal RNA genes of pea. *Plant Mol. Biol.* 8: 3-12.
- Kaiser, K., Murray, N.E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. in *DNA Cloning: Volume 1: A practical approach*, D.M. Glover, ed., IRL Press, Washington DC, p. 1-47.
- Keegstra, K. (1989). Transport and routing of proteins into chloroplasts. *Cell* 56: 247-253.
- Keegstra, K., Olsen, L.J. (1989). Chloroplast precursors and their transport across the envelope membranes. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 40: 471-501.
- Kemble, R.J., Mans, R.J., Gabay-Laughnan, S., Laughnan, J.R. (1983). Sequences homologous to episomal mitochondrial DNAs in the maize nuclear genome. *Nature* 304: 744-747.
- King, T.C., Sirdeskmukh, R., Schlessinger, D. (1986). Nucleolytic processing of ribonucleic acid transcripts in prokaryotes. *Microbiol. Rev.* 50: 428-451.
- Kirk, J.T.O., Tinley-Basset, R.A.E. (1978). *The Plastids*, Elsevier/North Holland Biomedical Press, Amsterdam.
- Koch, W., Edwards, K., Kössel, H. (1981). Sequencing of the 16S-23S spacer in a ribosomal RNA operon of Zea mays chloroplast DNA reveals two split tRNA genes. *Cell* 25: 203-213.

- Kohchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H., Ohya, K., (1988). Structure and organization of Marchantia polymorpha chloroplast genome. IV. Inverted repeat and small single copy regions. *J. Mol. Biol.* 203: 353-372.
- Kuhlemeier, C., Strittmatter, G., Ward, K., Chua, N.-H. (1989). The pea rbcS-3A promoter mediates light responsiveness but not organ specificity. *The Plant Cell* 1: 471-478.
- Kumano, M., Tomioka, N., Shizozaki, K., Sugiura, M. (1986). Analysis of the promoter region in the *rrn* operon from a blue-green alga, Anacystis nidulans. *Mol. Gen. Genet.* 202: 173-178.
- Kumano, M., Tomioka, N., Sugiura, M. (1983). The complete nucleotide sequence of a 23S rRNA gene from a blue-green alga Anacystis nidulans. *Gene* 24: 219-225.
- Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA* 82: 6955-6959.
- Lemieux, C., Boulanger, J., Otis, C., Turmel, M. (1989). Nucleotide sequence of the chloroplast large subunit rRNA gene from Chlamydomonas reinhardtii. *Nucl. Acids Res.* 17: 7997
- Levings III, C.S., Brown, G.G. (1989). Molecular biology of plant mitochondria. *Cell* 56: 171-179.
- Link, G. (1988). Photocontrol of plastid gene expression. *Plant Cell Environ.* 11: 329-338.
- Logemann, J., Schell, J., Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissue. *Anal. Biochem.* 168: 16-20.
- Lonsdale, D.M., Hodge, T.P., Howe, C.j., Stern, D.B. (1983). Maize mitochondrial DNA contains a sequence homologous to the ribulose-1,5-bisphosphate carboxylase large subunit gene of chloroplast DNA. *Cell* 34: 1007-1014.
- Mackay, J., Séguin, A., Lalonde, M. (1988). Genetic transformation of 9 *in vitro* clones of Alnus and Betula by Agrobacterium tumefaciens. *Plant Cell Reports* 7: 229-232.

- Maniatis, T., Fritsch, E.F., Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Mc Cown, B. H., Ellis, D. D. (1989). Unusual inheritance patterns of organelle DNA in conifers. *TIBTECH* 7: 257-258.
- Mc Garvey, P., Helling, R.B. (1989). Processing of chloroplast ribosomal RNA transcripts in *Euglena gracilis bacillaris*. *Curr. Genet.* 15: 363-370.
- Messing, J. (1983). New M13 vectors for cloning. *Meth. Enzymol.* 101: 20-78.
- Michel, F., Dujon, B. (1983). Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear encoded members. *EMBO J.* 2: 33-38.
- Michel, F., Umesono, K., Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns-a review. *Gene* 82: 5-30.
- Miyamura, S. Nagata, F., Kuroowa, T. (1986). Quantitative fluorescence microscopy on dynamic changes of plastid nucleoids during wheat development. *Protoplasma* 133: 66-72.
- Moon, E., Kao, T.-H., Wu, R. (1988). Rice mitochondrial genome contains a rearranged chloroplast gene cluster. *Mol. Gen. Genet.* 213: 247-253.
- Morelle, G. (1989). A plasmid extraction procedure on a miniprep scale. *BRL Focus* 11(1): 7-8.
- Morgens. P.H., Grabau, E.A., Gesteland, R.F. (1984). A novel soybean mitochondrial transcript resulting from a DNA rearrangement involving the 5S rRNA gene. *Nucl. Acid Res.* 12: 5665-5684.
- Mullet, J. E. (1988). Chloroplast development and gene expression. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 39: 475-502.
- Mullet, J.E., Klein, R.R. (1987). Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. *EMBO J.* 6: 571-579.

- Neale, D.B., Marshall, K.A., Sederoff, R.R. (1989). Chloroplast and mitochondrial DNA are paternally inherited in Sequoia sempervirens D. Don Endl. Proc. Natl. Acad. Sci. USA 86: 9347-9349.
- Nelson, M., Mc Clelland, M. (1989). Effect of site-specific methylation on DNA modifying methyltransferases and restriction endonucleases. Nucl. Acids Res. Suppl. 17: r389-r415.
- Newton, K. (1988). Plant mitochondrial genomes: organization, expression and variation. Ann. Rev. Plant Physiol. Plant Mol. Biol. 39: 503-532.
- Newcomb, W., Wood, S.M. (1987). Morphogenesis and fine structure of Frankia (Actinomycetales): the microsymbiont of nitrogen-fixing actinorhizal root nodules. Int. Rev. Cytol. 109: 1-88.
- Noller, H.F., Stern, S., Moazed, D., Powers, T. Svensson, P., Changchien, L.M. (1987). Studies on the architecture and function of 16S rRNA. Cold Spring Harbor Symposia on Quantitative Biology 52: 695-708.
- Normand, P., Bousquet, J. (1989). Phylogeny of nitrogenase sequences in Frankia and other nitrogen-fixing microorganisms. J. Mol. Evol. 29: 436-447.
- Normand, P., Lalonde, M. (1986). The genetics of actinorhizal Frankia: a review. Plant and Soil 90: 429-453.
- Normand, P., Simonet, P., Bardin, R. (1988). Conservation of nif sequences in Frankia. Mol. Gen. Genet. 213: 238-246.
- Normand, P., Simonet, P., Butour, J.L., Rosenberg, C., Moiroud, A, Lalonde, M. (1983). Plasmids in Frankia. J. Bacteriol. 155: 32-35.
- Ohshima, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-i., Inokuchi, H., Ozeki, H. (1988). Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification. J. Mol. Biol. 203: 281-298.

- Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-i., Inokuchi, H., Ozeki, H. (1986a). Complete nucleotide sequence of liverwort Marchantia polymorpha chloroplast DNA. *Plant Mol. Biol. Rep.* 4: 148-175.
- Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-i., Inokuchi, H., Ozeki, H. (1986b). Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature* 322: 572-574.
- Ozeki, H., Umesono, K., Inokuchi, H., Kohchi, T., Ohyama, K. (1989). The chloroplast genome of plants: a unique origin. *Genome* 31: 169-174.
- Pace, N.R., Olsen, G.J., Woese, C.R. (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* 45: 325-326.
- Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W., Manhart, J. R. (1988a). Chloroplast DNA variation and plant phylogeny. *Ann. Missouri Bot. Gard.* 75: 1180-1205.
- Palmer, P.D., Osorio, B., Thompson, W.F. (1988b). Evolutionary significance of inversions in legume chloroplast DNAs. *Curr. Genet.* 14: 65-74.
- Palmer, J.D., Stein, D.B., (1986). Conservation of chloroplast genome structure among vascular plants. *Curr. Genet.* 10: 823-833.
- Palmer, J.D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19: 325-354.
- Piechulla, B., Pichersky, E., Cashmore, A.R., Grissem, W. (1989). Expression of nuclear and plastid genes for photosynthesis-specific proteins during tomato fruit development and ripening. *Plant Mol. Biol.* 7: 367-367.
- Raleigh, E.A. (1987). Restriction and modification in vivo by Escherichia coli K12. *Meth. Enzymol.* 152: 130-141.
- Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J., Hanahan, D. (1988). McrA and McrB restriction phenotypes of some E. coli strains and implications for gene cloning. *Nucl. Acid Res.* 16: 1563-1575.

- Raleigh, E.A., Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA: 83: 9070-9074.
- Raué, H.A., Klootwijk, J., Musters, W. (1988). Evolutionary conservation of structure and function of high molecular weight ribosomal RNA. Prog. Biophys. molec. Biol. 51: 77-129.
- Rochaix, J.D. (1981). Organization, function, expression of the chloroplast DNA of Chlamydomonas reinhardtii. Experientia 37: 323-332.
- Rochaix, J.D., Malnoe, P. (1978). Anatomy of the chloroplast ribosomal DNA of Chlamydomonas reinhardtii. Cell 15: 661-670.
- Rochaix, J.D., Rahire, M. Michel, M. (1985). The chloroplast ribosomal intron of Chlamydomonas reinhardtii codes for a polypeptide related to mitochondrial maturases. Nucl. Acids Res. 13: 975-984.
- Sanger, F., Niklen, S., Coulson, A.R. (1977). DNA Sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.
- Schaal, B.A., Learn Jr., G.H. (1988). Ribosomal DNA variation within and among plant populations. Ann. Missouri Bot. Gard. 75: 1207-1216.
- Schaffer, H.E., Sederoff, R.R. (1981). Estimation of DNA fragment size from agarose gels. Anal. Biochem. 115: 113-122.
- Schwarz, Z., Kössel, H. (1980). The primary structure of 16S rDNA from Zea mays chloroplast is homologous to E. coli 16S rRNA. Nature 283: 739-742.
- Schwarz, Z., Kössel, H., Schwarz, E., Bogorad, L. (1981). A gene coding for tRNA^{Val} is located near 5' terminus of 16S rRNA gene in Zea mays chloroplast genome. Proc. Natl. Acad. Sci. USA: 78: 4748-4752.
- Scott, N.S., Timmis, J.N. (1984). Homologies between nuclear and plastid DNA in spinach. Theor. Appl. Genet. 67: 279-288.
- Shine J., Dalgarno, L. (1974). The 3' terminal sequences of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71: 1342-1346.

- Siebert, P.D., Bluford, P., Fukuda, M. (1989). mRNA and gene structure determination with synthetic oligonucleotide probes. *Clonetechniques: A roadmap for the molecular biologist 1*: 6-9.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, C., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. (1986a). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5: 2043-2049.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. (1986b). The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Mol. Biol. Rep.* 4: 110-147.
- Sibbald, P.R. (1988). Patterns of base usage, nearest neighbour analysis and identification of genes in two completely sequenced chloroplast genomes. *Curr. Genet.* 13: 523-530.
- Simonet, P., Haurat, J., Normand, P., Bardin, R., Moiroud, A. (1986). Localization of *nif* genes on a large plasmid in *Frankia* sp. strain ULQ0132105009. *Mol. Gen. Genet.* 204: 492-495.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Steinmetz, A., Weil, J.-H. (1989). Protein synthesis in chloroplasts. *The Biochemistry of Plants.* 15: 193-224.
- Stern, D.B., Gruissem, W. (1989). Chloroplast mRNA 3' end maturation is biochemically distinct from prokaryotic mRNA processing. *Plant Mol. Biol.* 13: 615-625.
- Stern, D.B., Jones, H., Gruissem, W. (1989). Function of plastid mRNA 3' inverted repeats: RNA stabilization and gene-specific protein binding. *J. Biol. Chem.* 264: 18742-18750.

- Stern, D.B., Gruissem, W. (1987). Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* 51: 1145-1157.
- Stern, D.B., Hodge, T.P., Lonsdale, D.M. (1984). Homology between the ribosomal DNA of *Escherichia coli* and mitochondrial DNA preparations of maize is principally to sequences other than mitochondrial rRNA genes. *Plant Mol. Biol.* 3: 355-361.
- Stern, D.B., Palmer, J.D. (1984). Extensive and widespread homologies between mitochondrial DNA and chloroplast DNA in plants. *Proc. Natl. Acad. Sci. USA* 81: 1946-1950.
- Stern, D.B., Lonsdale, D.M. (1982). Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* 229: 698-702.
- Strauss, S.H., Palmer, J.D., Howe, G.T., Doerksen, A.H. (1988). Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proc. Natl. Acad. Sci. USA* 85: 3898-3902.
- Strittmatter, G., Gozdzicka-Jozefiak, A., Kössel, H. (1985). Identification of an rRNA operon promoter from *Zea mays* chloroplasts which excludes the proximal tRNA^{Val} (GAC) from the primary transcript. *EMBO J.* 4: 599-604.
- Strittmatter, G., Kössel, H. (1984). Cotranscription and processing of 23S, 4.5S and 5S in chloroplasts from *Zea mays*. *Nucl. Acids Res.* 12: 7633-7647.
- Sugiura, M. (1989a). The Chloroplast Genome. *The Biochemistry of Plants.* 15: 133-149.
- Sugiura, M. (1989b). The chloroplast chromosomes in land plants. *Annu. Rev. Cell. Biol.* 5: 51-70.
- Sugiura, M., Shinozaki, K., Tanaka, M., Hayashida, M., Wakasugi, T., Matsubayashi, T., Ohto, C., Torazawa, K., Meng, B.Y., Hidaka, T., Zaita, N. (1987). Split genes and CIS/TRANS splicing in tobacco chloroplasts. in *Plant Molecular Biology*, D. Von Wettstein and N.-H. Chua eds., Proceedings of NATO Advanced Study, Institute on Plant Molecular Biology, June 10-19, 1987 at the Carlsberg Laboratory, Copenhagen, Denmark, Plenum Press, New York. p 65-76.

- Tahar, S.B., Bottomley, W., Whitfield, P.R. (1986). Characterization of the spinach chloroplast genes for the S4 ribosomal protein, tRNA^{Trp}(UGU) and tRNA^{Ser}(GGA). *Plant Mol. Biol.* 76: 63-70.
- Takaiwa, F., Sugiura, M. (1982). The complete nucleotide sequence of a 23-S rRNA gene from tobacco chloroplasts. *Eur. J. Biochem.* 124: 13-19.
- Timmis, J.N., Scott, N.S. (1983). Sequence homology between spinach nuclear and chloroplast genomes. *Nature* 305: 65-67.
- Tohdoh, N., Sugiura, M. (1982). The complete nucleotide sequence of a 16S ribosomal RNA gene from tobacco chloroplasts. *Gene* 17: 213-218.
- Tohdoh, N., Shinozaki, K., Sugiura, M. (1981). Sequence of a putative promoter region for the rRNA genes of tobacco chloroplast DNA. *Nucl. Acids Res.* 9: 5399-5406.
- Tomioka, N., Sugiura, M. (1984). Nucleotide sequence of the 16S-23S spacer region in the rRNA operon from a blue-green alga, Anacystis nidulans. *Mol. Gen. Genet.* 193: 427-430.
- Tomioka, N., Sugiura, M. (1983). The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, Anacystis nidulans. *Mol. Gen. Genet.* 191: 46-50.
- Torrey, J.G. (1983). Casuarina: Actinorhizal dinitrogen-fixing tree of the tropics. in *Casuarina Ecology, Management and Utilization*, Midgley, S.J., Turnbull, J.W., Johnston, R.D. eds., Canberra, Act, Australia, Commonwealth Scientific and Industrial Research Organization, pp. 193-204.
- Torrey, J.G. (1978). Nitrogen fixation by actinomycete-nodulated angiosperms. *Biosci.* 28: 586-592.
- Van den Eynde, H., De Baere, R., De Roeck, E., Van de Peer, Y., Vanderberghe, A., Willekens, P., De Wachter, R. (1988). *J. Mol. Evol.* 27: 126-132.
- Verma, D.P.S., Delauney, A.J. (1988). Root nodule symbiosis: nodulins and nodulin genes. Chapter 10 in *Temporal and spatial regulation of plant genes*, eds D.P. Verma and R.B. Goldberg, Springer-Verlag/Wien, pp. 169-199.

- Wallace, B., Miyada, G. (1987). Oligonucleotide probes for the screening of recombinant DNA libraries. *Meth. Enzymol.* 152: 432-442.
- Weisbeek, P., Hageman, J., Smeekens, S., de Boer, D., Cremers, F. (1987). Chloroplast-specific import and routing of proteins. in *Plant Molecular Biology*, D. Von Wettstein and N.-H. Chua eds., Proceedings of NATO Advanced Study, Institute on Plant Molecular Biology, June 10-19, 1987 at the Carlsberg Laboratory, Copenhagen, Denmark, Plenum Press, New York. p 77-91.
- Wheeler, C.T., Hooker, J.E., Crowe, A. Berrie, A.M.M. (1986). The improvement and utilization in forestry of nitrogen fixation by actinorhizal plants with special reference to Alnus in Scotland. *Plant and Soil* 90: 393-406.
- White, J.A., Williams, S.E. (1985). Symbiotic N₂-fixation by Frankia: actinorhizae establishment on disturbed lands. in *Proceedings: American Society for surface mining and reclamation* p.289-297.
- Whittaker, P.A., Campbell, A.J.B., Southern, E.M., Murray, N.E. (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a new X vector. *Nucl. Acids Res.* 16: 6725-6736.
- Wittmann, H.G. (1983). Architecture of prokaryotic ribosomes. *Annu. Rev. Biochem.* 52: 36-65.
- Woese, C.R., Gutell, R.R. (1989). Evidence for several higher order structural elements in ribosomal RNA. *Proc. Natl. Acad. Sci. USA.* 86: 3119-3122.
- Woese, C.R., Gutell, R., Gupta, R., Noller, H.F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* 47: 621-669.
- Wolfe, K. H., Sharp, P. M. (1988). Identification of functional open reading frames in chloroplast genomes. *Gene* 66: 215-222.
- Wood, W.B. (1966). Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.* 16: 118-133.

- Yamada, T., Shimaji, M. (1987). An intron in the 23S rRNA gene of the Chlorella chloroplasts: Complete nucleotide sequence of the 23S rRNA gene. *Curr. Genet.* 11: 347-352.
- Yamada, T., Shimaji, M. (1986). Peculiar feature of the organization of rRNA genes of the Chlorella chloroplast DNA. *Nucl. Acid Res.* 14: 3827-3839.
- Yamaguchi-Shinozaki, K., Shinozaki, K., Sugiura, M. (1987). Processing of precursor tRNAs in a chloroplast lysate: processing of the 5'-end involves endonucleolytic cleavage by an RNase P-like enzyme and precedes 3'-end maturation. *FEBS Lett.* 215: 132-136.
- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., Woese, C.R. (1985). Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* 82: 4443-4447.

APPENDIX 1**WHEN NODULINS BECOME RIBOSOMAL RNA GENES**

The nature of the genes selected by hybridization had been thought to be nodulin genes but clues along the way were slowly suggesting otherwise.

Southern analysis of total alder genomic DNA hybridized to each of the cDNA probes pAnod4, pAnod5 (and pAnod2) suggested that the clones were independent and of low copy number as the fragments that hybridized to probes pAnod4 and pAnod5 were distinct and the pattern observed was simple. Upon screening of the alder genomic lambda library, however, the frequencies of appearance of the positive clones was estimated to be ~0.05% which seemed quite high for the expected frequency of a single copy gene. The numbers suggest a multicopy gene or one for which the DNA was over represented in some way in the library. By this time, the partial sequences for the cDNA inserts had been compared to data banks using the MicroGenie software package and no homologous sequences were detected to any plant or organelle DNA gene present in the data banks.

At the same time in anticipation of transcription studies, total RNA was being isolated from root and nodule

tissue from alder and analyzed by electrophoresis. The sizes of the hybridizing bands previously detected using poly A+ RNA and nick-translated plasmids as probes, were reproduced using total nodule RNA and probes derived from the cDNA inserts. This demonstrated that the labeled plasmid had detected no homologous sequences on Northern transfers and the sizes of the RNA species detected were reproducible. More importantly, however, positive signals were detected for the *E. coli* rRNA markers on those Northern blots which contained this marker. Because of the sizes of the mRNAs detected and the negative results from data bank searches of plant/organelle DNA/RNA sequences using portions of sequenced cDNAs, ribosomal DNAs had been previously excluded.

Later comparisons to data banks using the NRC Wisconsin package demonstrated that the cDNA sequences were ribosomal in origin. pAnod4 and pAnod5 probe inserts showed a high degree of homology to bacterial 23S and 16S ribosomal RNA genes, whereas pAnod2 probe insert was found to be homologous to nuclear 18S ribosomal RNA genes. It was determined that the data comparison programs used previously were defective.

Given the identity of the genes as ribosomal RNA genes, then it is expected that the pattern of expression observed

is similar because the 16S and 23S ribosomal RNA genes are cotranscribed (Gruissem, 1989b).

Data obtained from the nucleotide sequence of subclones of the lambda clones selected using the probes pAnod4 and pAnod5 are unequivocally organellar ribosomal RNA species by comparisons to data banks using the IBI and Microgenie programs. λ Anod4-1 and λ Anod4-2 sequences show homologies to chloroplast 23S and 16S ribosomal DNA and λ Anod5-5, to mitochondrial 18S ribosomal DNA.

Two questions arise then from this exercise: first why did bacterial sequences select chloroplast and mitochondrial sequences, secondly why was chloroplast DNA preferentially isolated from the alder genomic library. According to sequence analysis, the DNA sequence of the cDNA inserts for pAnod4 and pAnod5 show a greater homology to bacterial ribosomal DNA. Under the hybridization conditions used in the screening of the library and the subsequent Southern transfer analyses, it is possible that the probes detect homologous sequences in the organellar genomes (Stern *et al.*, 1984).

It is possible that chloroplast genes were cloned from alder total genomic DNA as whole leaf tissue was used for DNA isolation. Palmer (Palmer *et al.*, 1988a) suggests the

use of total DNA as opposed to chloroplast DNA to study chloroplast DNA variation because "total DNA extraction efficiencies are nearly quantitative and provide 5-100 times higher yields of cpDNA than cpDNA extraction procedures". Chloroplast DNA is present in several thousand copies in the typical leaf cell (Bendich, 1987). Had nuclei been isolated prior to the preparation of DNA then chloroplast contamination may have been reduced, and possibly nuclear ribosomal genes would have been detected instead as bacterial rDNA sequences cross hybridize to chloroplast, mitochondrial and nuclear genes (see discussion).

Therefore it makes sense that the genes selected with the probes pAnod4 and pAnod5 are linked and cotranscribed as are ribosomal RNA genes, and this also explains the similar patterns of expression which had been observed previously (T. Gleeson, M.Sc. thesis).

APPENDIX 2

CHARACTERIZATION OF THE CLONE λ Anod5-5

The partial characterization of the clone λ Anod5-5 is described here. The origin of this clone was outlined in Results section IV.E.2. By computer analysis it shows homology to mitochondrial 18S ribosomal sequences.

I. RESTRICTION ANALYSIS

The size of the fragments in clone λ Anod5-5 (Figure 14) hybridizing to the pAnod5 probe (Figure 13) are not the expected sizes as calculated from the alder genomic Southern blots (Figure 4). There can be at least two possible explanations for this discrepancy. The first, as previously mentioned for λ Anod4-1, is that the restriction sites which flank the hybridizing sequence were not cloned in the lambda recombinant. The second possibility is that, unlike chloroplast DNA which is very well represented in total DNA preparations (Palmer *et al.*, 1988), mitochondrial DNA may be under represented and the hybridizing bands may have been masked by the chloroplast bands. It is noteworthy that the BamHI fragment hybridizing to the pAnod5 probe of both

XAnod4-1 and XAnod5-5 clones appear to be the same size if an error of 10% is taken into account.

II. SEQUENCE COMPARISON TO MITOCHONDRIAL 18S RIBOSOMAL GENES

From the λ Anod5-5 clone, a BglII fragment of 2.5 kb was cloned into M3mp19 and a partial sequence was obtained (Figure 15). Homology of the 305 bp sequence as determined for the left end of this BglII fragment is greater to plant mitochondrial 18S rDNA than it is to chloroplast 16S rDNA. The BglII fragment also shows homology to other 18S genes sequenced ie. maize, wheat (the scores for homology were similar as defined by the IBI Pustell Sequence Analysis Programs). The percent homologies detected for this alder sequence with respect to the soybean, maize (see Figure 15) for sequence alignments) and wheat 18S mitochondrial rDNA genes were 96.3%, 96.5%, 96.6%, respectively, in contrast with those of tobacco and liverwort 16S chloroplast rDNA which were 68.3%, and 70% respectively. Localization of the MBg50 subclone sequence on the soybean mitochondrial sequence (Grabau, 1985) shows that the cloned portion is found in the 5' region of the gene (position 406-703, out of 1990 nucleotides). The nucleotides in this region include one "universal region" (U3), and part of another (U2) which is within the universal core of the 16S molecule defined by Woese (Woese *et al.*, 1983); these are regions of

conservation of nucleotide sequence between a number of organism, including E. coli. Thus we conclude that the clone represents alder mitochondrial 18S rDNA sequence but further analysis is needed to confirm this conclusion. Fragments of this clone could be used to detect mitochondrial genomes and RFLPs.

FIGURE 13: HYBRIDIZATION PATTERN OF λ Anod5-5 TO pAnod5
INSERT PROBE

An autoradiogram of a Southern transfer of λ Anod5-5 DNA. (1 μ g) aliquots of λ Anod5-5 DNA digested with a number of restriction enzymes was electrophoresed, blotted and hybridized to pAnod5 cDNA insert. Digests are: 1, BamHI; 2, BamHI+BglIII; 3, BamHI+EcoRI; 4, BamHI+HindIII; 5, BamHI+MspI; 6, BamHI+XhoI; 7, marker lambda DNA + HindIII and ϕ X174 + HaeIII; 8, EcoRI; 9, EcoRI+BglIII; 10, EcoRI+HindIII; 11, EcoRI+MspI; 12, EcoRI+XhoI; 13, HindIII. Sizes (in kb) of the molecular weight standards of lane 7 are indicated to the left of the panel. The arrow indicates the origin.

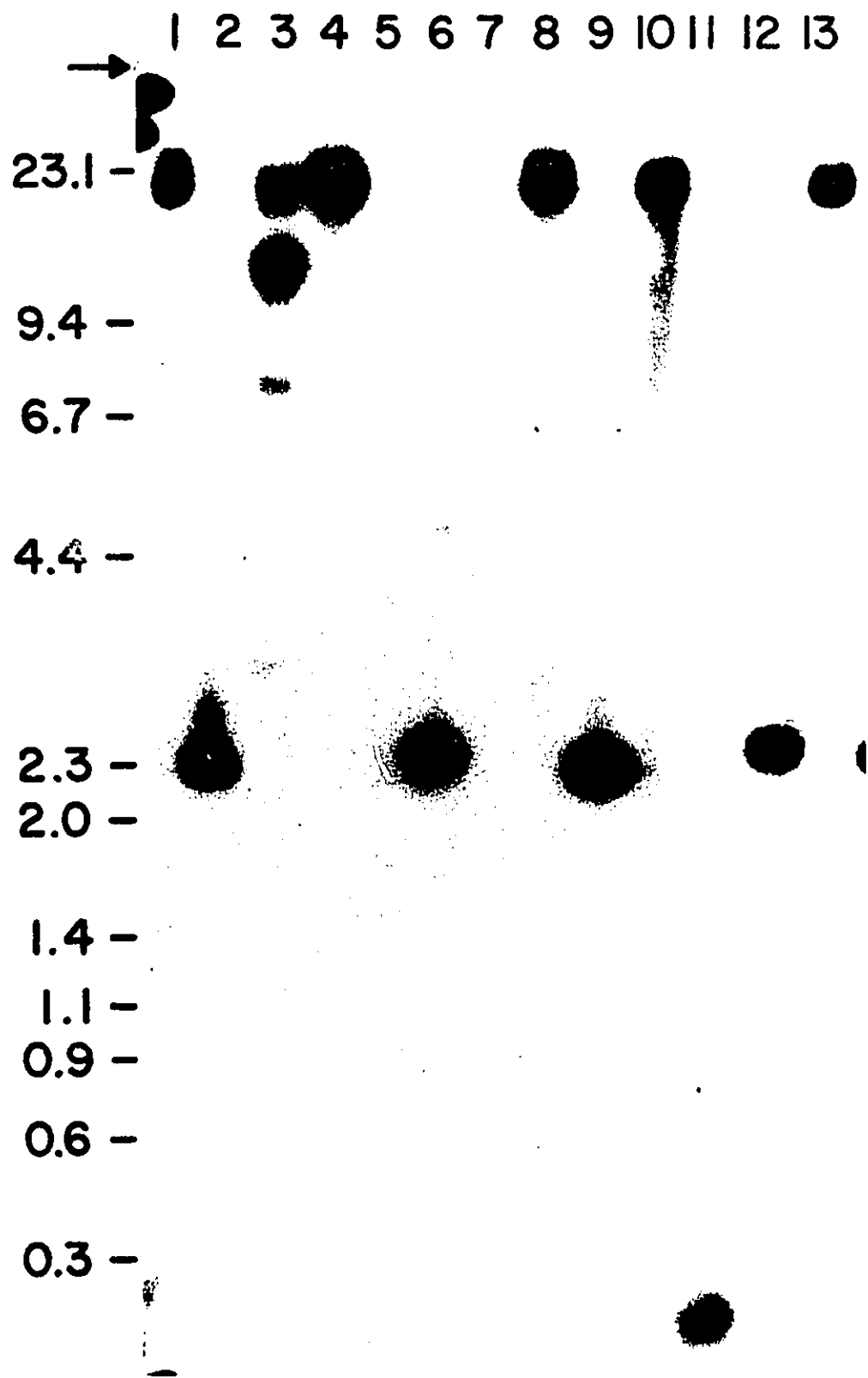
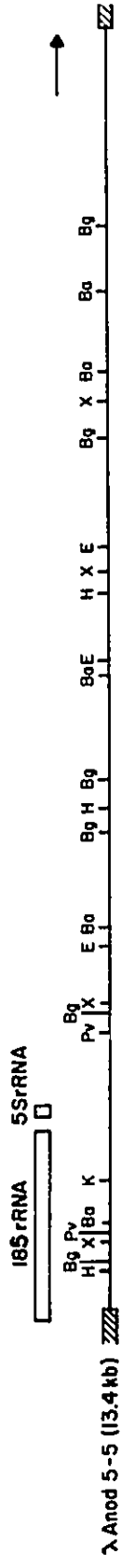


FIGURE 14: RESTRICTION MAP OF THE λ Anod5-5 CLONE

The clone λ Anod5-5 is estimated to contain 13.4 kb of DNA (by restriction analysis). Restriction sites are as follows: Ba, BamHI; Bg, BglII; E, EcoRI; H, HindIII; K, KpnI; Pv, PvuII; X, XhoI. The position of the sites was determined by restriction analysis and partial sequencing of the leftmost 2.5 kb BglII fragment (MBg50 and Bg5K, subclones). Bar is 1 kb.



1 kb

Bgl

Subclones

FIGURE 15: NUCLEOTIDE SEQUENCE OF SUBCLONE MBg50 AND
SEQUENCE COMPARISON TO THE MAIZE MITOCHONDRIAL 18S rRNA GENE

The sequence is derived from the left end of the clone, as oriented in Figure 14. Restriction enzyme sites are: Hp, HpaII; M, MboI; X, XhoI. The BglII fragment was cloned into a BamHI site thus both BglII and BamHI sites (position 305) were lost.

The maize mitochondrial 18S rRNA gene nucleotide sequence (Chao *et al.*, 1984) is included on the second line of the MBg50 sequence. Dots indicate the same nucleotides, dashes indicate deletions and (^) show insertions. The maize sequence is taken from the GenBank file MZEMTGE2 and begins at position 1019.

M
 TCGAAGCCGA TCCAGCAATA TCCGCGTGAG GTAAGAAGGC AATGCGCTTG TAAAGCTCTT 60
 G. λ_a λ_c - TG. λ_g λ_c

M X Hp 120
 TCGTCAGTGC GCGGATCATG ACAGACTCGA GGAAGAAGCC CCGGCTAACT CCGTGCCAGC
 λ_g - λ_g

180
 AGCCGCGGTA AGACGGGGGG GGCAAGTGTT CTTCGGAATG ACTGGGCGTA AAGGGCACCG
 -

240
 TAGGCGGTGA ATCGGGTTGA AAGTTAAAGT CGCCAAAAAC TGGTGGAATG CTCTCGAAAC
 G G ... C

Hp300
 CAATCACTT GAGTGAAGAG CAGAGGAGAG TGAATTTTCG TGTGTAGGGG TGAAATCCGG
 - - T

M 305
 AGATC

APPENDIX 3

ALDER AND BIRCH SOUTHERN TRANSFERS: COMPARISON

Birch and alders are closely related taxonomically (Normand and Lalonde, 1986) ie. both belong to the same family, the Betulaceae, but only alder forms symbiotic associations with Frankia. A future comparative study of their genes was envisaged to determine if birch had the genetic makeup for nodulation. Given the change in nature of the cDNAs, information gathered from birch can be used in a predictive sense as to the relationship between alder and birch. In order to start the characterization of birch genes, Southern blots of birch genomic DNA isolated from B. papyrifera were also prepared and screened with the same cDNA probes previously used for alder (Figure 16). The sizes of hybridizing fragments for birch DNA to the three cDNA inserts are indicated in Table 5 and can be compared to those found in alder for the same enzymes in Table 5.

Although there is variation between alder and birch in the size of fragments detected for the same restriction enzyme with the same probe, the overall pattern of hybridization is well conserved. The fragment sizes observed for both angiosperms are virtually the same given the estimated error ($\sim\pm 10\%$). The birch BamHI fragment hybridizing to the pAnod4 probe is 2.9 kb in length as

opposed to that of 3.0 kb in alder. Based on the similarities of hybridizing restriction fragment sizes, it is proposed that both birch and alder would have the same genomic structure for the rRNA genes. In order to determine whether the restriction fragment lengths for alder and birch are identical, it would have been best to electrophorese digests of both alder and birch genomic DNAs on the same gel and subject the gel to Southern transfer and hybridization to the probes. This was attempted but was unsuccessful as the alder DNA prepared for this analysis resisted agarose gel electrophoresis.

The differences observed in the sizes of the fragments hybridizing could result from RFLP (restriction fragment length polymorphisms) due to mutation or differential methylation. Base changes are estimated to be few in birch as compared to alder because the homologies of the alder sequences to tobacco are ~98% and these two groups of plants are further removed. Alternatively, the enzymes BamHI, HpaII, MspI, Sau3A, used in this study are sensitive to methylation of the cytosine residue within their recognition site (Nelson and McClelland, 1989). As plant DNAs are known to be highly methylated (Belanger and Hepburn, 1990), and these authors noted especially that ribosomal DNA is heavily methylated and show no shortage of CpG sequences, perhaps the extent of methylation of the cytosine residues between

alder and birch is not the same; but this is conjecture at this time. It is noteworthy that most of the changes in fragment sizes occur for the pAnod2 probe; as this probe was homologous to nuclear ribosomal RNA genes, then perhaps methylation of the nuclear DNA will be more relevant.

Then, the relatedness observed between alder and birch based on "phenotypic" traits will more than likely be concordant with the molecular data, especially if rRNA genes were to be isolated and sequenced from birch. These preliminary studies indicate that there are similarities of structure and sequence between these two angiosperms and probes from alder would potentially be useful for the isolation of birch genes.

FIGURE 16: HYBRIDIZATION OF BIRCH DNA TO cDNA INSERT PROBES

A Southern transfer of birch genomic DNA was screened sequentially with cDNA inserts derived from A, pAnod2; B, pAnod4 and C, pAnod5. Five micrograms of DNA was digested with 1, BamHI; 2, BstI; 3, HapII; 4, MspI; 5, Sau3A; 6, MboI; 7, EcoRI. The arrow indicates the origin. Sizes of molecular weight standards (in kb) on the left of the panels are those of lambda DNA digested with HindIII and ϕ X174 DNA digested with HaeIII.

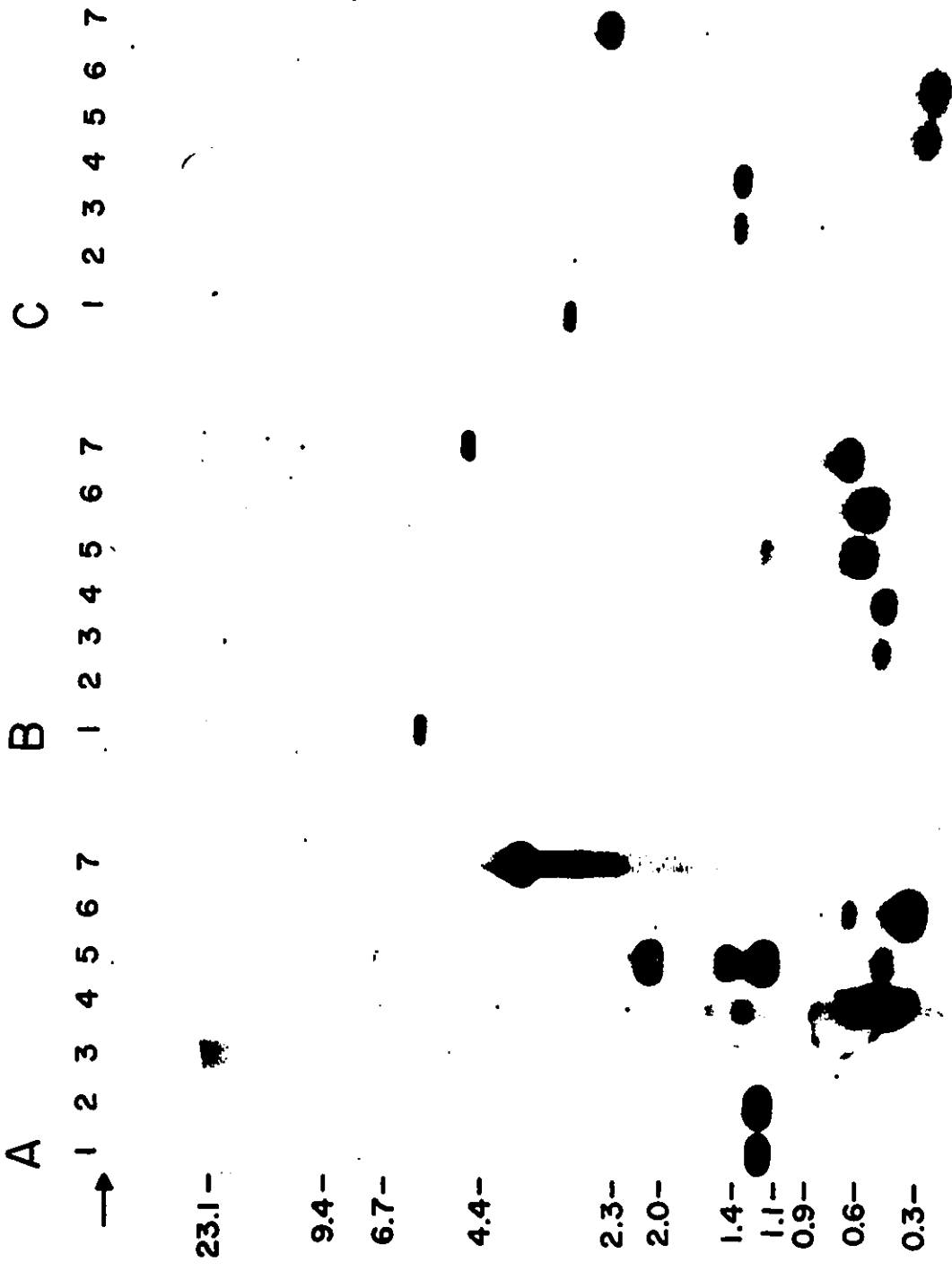


TABLE 5: FRAGMENT SIZES OF ALDER AND BIRCH TOTAL DNA
HYBRIDIZING TO EACH cDNA INSERT PROBE

Presented are the calculated fragment sizes observed for the alder and birch total DNA Southern blots, taken from four independent blots for alder (A), and three independent blots for birch (B). Sizes are in kb. The error is estimated to be +/- 10%.

RESTRICTION ENZYME	PROBES					
	pAnod2		pAnod4		pAnod5	
	A	B	A	B	A	B
BamHI	5.5 1.1	1.2	5.8	5.8	3.0	2.9
BglIII	~25.0 ~23.0 15.1 7.6	-	5.1	-	3.8	-
BstI	-	1.2	-	5.8 5.5 4.7 4.4 3.8 3.4 2.9	-	2.9
EcoRI	3.8	3.5	4.7 0.7	4.6 0.6	2.6	2.3
HapII	-	~25.0	-	0.7 0.5 0.3	-	1.3
HindIII	7.6	-	1.3 1.1	-	7.9	-
MboI	0.3 0.2	0.6 0.3 0.2	0.5	0.5	0.3	0.3
MspI	1.6 1.3 1.1 0.8 0.7 0.5 0.3	1.6 1.4 1.2 0.9 0.6 0.4	0.7 0.5 0.3	0.7 0.5 0.3	1.4	1.3
Sau3A	-	2.0 1.4 1.1 0.5	-	1.1 1.0 0.6	-	0.3