PLASTID DNA SEQUENCE HOMOLOGIES WITHIN THE NUCLEAR GENOMES OF HIGHER PLANT SPECIES.

by

MICHAEL A. AYLIFFE B.Sc.(Hons.)

Department of Genetics, University of Adelaide.

A thesis submitted to the University of Adelaide for the degree of Doctor of Philosophy.

April, 1992.

TABLE OF CONTENTS

page

Title page............ i

Table of contents............ ii

Abstract............ vi

Declaration............ viii

Acknowledgements............ ix

Abbreviations............ x

Publications............ xi

Chapter 1: Introduction............ 1

1.1: The nucleus............ 1

1.2: The mitochondrion............ 1

1.3: The plastid............ 5

1.4: Nuclear dependence of cytoplasmic organelles............ 7

1.5: Origin of the mitochondrion and plastid : the endosymbiont theory.. 8

1.6: Transfer of genetic information between organelles............ 9

1.7: Promiscuous DNA sequences: plastid sequences present within the mitochondrial genome............ 12

1.8: Nuclear DNA sequence homologies in the mitochondrial genome.... 16

1.9: Mitochondrial DNA sequence homologies in the nuclear genome.... 17

1.10: Plastid DNA sequence homologies in the nuclear genome............ 21

1.11: Promiscuous sequences present within the plastid genome? 23

1.12: Mechanisms of promiscuous DNA transfer............ 26

1.13: Expression of promiscuous sequences............ 29

1.14: Summary............ 31

Chapter 2: Materials and Methods.. 32

2.1: Materials.. 32

2.2: Stock solutions.. 33

2.3: Bacterial growth media.. 34

2.4: *E. coli* and bacteriophage strains... 34

2.5: Isolation of plasmid and bacteriophage DNA.................................. 34

2.6: Isolation of plant nucleic acids.. 36

2.7: Recombinant DNA techniques... 38

2.8: Electrophoresis.. 38

2.9: Nucleic acid membrane transfer and hybridization conditions......... 39

2.10: Double stranded plasmid DNA sequencing.................................. 41

2.11: Oligonucleotide primer synthesis.. 42

2.12: Polymerase chain reaction.. 42

Chapter 3: Plastid DNA sequence homologies within higher plant nuclear genomes.. 43

3.1: Species with a constant nuclear hybridization pattern to plastid DNA probes.. 45

3.2: Species which show variation in nuclear hybridization pattern to plastid DNA probes.. 46

3.3: Individual plant variation in nuclear hybridization patterns to plastid DNA probes.. 47

3.4: Hypervariability of plastid DNA sequence homologies in the nuclear genome of *Beta vulgaris*.. 48

3.5: Breeding systems.. 50

3.6: Somatic variation of nuclear sequences with plastid DNA homology.. 51

3.7: Relationship between genome size and presence of plastid DNA sequence homologies.. 53

3.8: Summary 54

Chapter 4: Promiscuous plastid sequences in the tobacco nuclear genome 55

4.1: Quantification of plastid DNA homologies in tobacco nuclear DNA 57

4.2: Nuclear origin of these plastid DNA homologies 59

4.3: Construction of a tobacco nuclear genomic library 61

4.4: Plating of the genomic library 61

4.5: Screening of the genomic library 63

4.6: Summary 63

Chapter 5: Characterization of lambda clones 3D, 3b, 3-2, 1H and 2d 65

5.1: Lambda clone 3D 66

5.2: Lambda clone 3b 70

5.3: Lambda clone 3-2 72

5.4: Regions of sequence homology between lambda clones 72

5.5: Lambda clones 1H and 2d 74

5.6: Summary 74

Chapter 6: Lambda clone 1M 76

6.1: Genomic location of clone 1M 77

6.2: Deletions in lambda clones 1M and 3D 78

6.3: Summary 80

Chapter 7: Transcription of promiscuous plastid sequences in the plant nuclear genome 82

7.1: Northern blot analysis 83

7.2: Tobacco cDNA library.. 83

7.3: cDNA clone with homology to the 7.7 kbp PstI DNA fragment....... 84

7.4: Origin of this cDNA clone.. 85

7.5: Summary.. 86

Chapter 8: Discussion.. 88

References.. 94

108

* All figures occur at the end of each chapter.

Abstract

Within the plant cell are three distinct genetic compartments: the nucleus, the plastid and the mitochondrion. DNA sequences that are common to more than one of these organelle genomes, termed promiscuous DNA sequences (Ellis, 1982), are believed to be the result of transfer of genetic information from one organelle and subsequent integration into the genome of another. These sequences have been identified in both the nuclear and mitochondrial genome but never satisfactorily in the plastid genome. The mechanism of transfer of these sequences is yet to be elucidated. The aim of this study was to further characterize plastid DNA sequence homologies within higher plant nuclear genomes.

Southern blot analysis identified plastid DNA sequence homologies in the nuclear genomes of all plant species examined, except for *Arabidopsis thaliana.* As assessed by restriction fragment variability, in some plant species a homogeneous arrangement of these sequences is observed while other species show different amounts of intraspecific variation. This sequence variability may be attributable to the different reproductive mechanisms employed by each species or alternatively reflect the diversity of the gene pool from which each cultivar was derived. Plastid DNA sequence homologies in the nuclear genome of *Beta vulgaris* were the most heterogeneous being potentially somatically unstable. Southern hybridization suggested the presence of very large, contiguous tracts of plastid DNA in the *Nicotiana tabacum* nuclear genome, in a high copy number (57 copies per 1c nucleus) when compared with other species (Steele Scott and Timmis, 1984; Pichersky *et al*, 1991: Pichersky and Tanksley, 1988).

Isolated from a tobacco nuclear genomic library were 300 lambda clones that showed homology to three plastid DNA probes. Six of these clones were examined in varying amounts of detail and five were shown to be greater than 99% homologous in their entirety to the plastid genome. These clones were not of *bona fide* plastid origin as they did not contain functional open reading frames in essential photosynthetic genes. It was therefore concluded that integrated within the tobacco nuclear genome are multiple copies of large (ie. in excess of 18 kbp), contiguous tracts of plastid DNA.

These five clones encompassed 30% of the entire tobacco plastid genome and thus encoded nuclear equivalents of many plastid genes. Quantitative PCR analysis of one of these sequences gave an estimate of 15 $\pm$5 copies per 1c tobacco nuclear genome. As all 15 copies of this sequence contained the same 41 bp deletion they are presumed to have arisen from duplications that occurred after a single transposition event. The apparent multiplication of these promiscuous sequences within the nuclear genome may account for the absence of any unifying sequence motifs at the integration sites of promiscuous nuclear sequences identified to date.

The remaining lambda clone contained both truly nuclear and plastid derived DNA sequences. These nuclear sequences showed homology to a moderately repetitive dispersed repeat present within the tobacco nuclear genome. The region of plastid DNA homology of this clone showed greater sequence divergence when compared with the true plastid genome than the five other lambda clones, suggesting that it represented a more ancient transposition of plastid DNA to the nucleus. Associated with nucleotide deletions and insertions in these lambda clones were sequence motifs suggesting that processes such as DNA replication slippage and palindrome mediated DNA excision were involved in the formation of these mutations. As the lambda clones encompass many regions of the plastid genome that are not transcribed, transfer of these sequences to the plant nucleus would not appear to have been mediated via an RNA intermediate, as has been suggested for other species (Schuster and Brennicke, 1987; Nugent and Palmer, 1991).

The presence of large tracts of plastid DNA in the tobacco nuclear genome contrasts the arrangement of such sequences in the nuclear genomes of other studied plant species (Steele Scott and Timmis, 1984; Pichersky *et al*, 1991; Pichersky and Tanksley, 1988).

Declaration

I declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and to the best of my knowledge and belief contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I give my consent for this thesis to be made available for photocopying and loan.

Michael A. Ayliffe

Acknowledgements

I wish to thank my supervisors, Dr J N Timmis and Dr N S Scott for their advice, support, encouragement and friendship throughout this study.

I am further indebted to the Cotton Seed Distributors Ltd. (Wee Waa, N.S.W.) for financial support in the form of a Postgraduate Research Scholarship and I applaud their commitment to Australian scientific research.

Many of the staff and students (both past and present) in the Genetics Department at the University of Adelaide have contributed at both a personal and professional level for which I am grateful. In particular I would like to acknowledge Dr M O'Connell, Dr J Kelly and Dr C Dowzer for their advice and help throughout the course of this study. In addition I thank Doug Pottrell for photographic and computing assistance, Robert Shroff for help with computing and proof-reading and Dr D Anderson for help with age estimations based on sequence divergence.

I also wish to thank Dr M Sugiura for providing the clone bank of the tobacco plastid genome, Dr D Woodcock for supplying *E. coli* strain NW2, Dr P R Whitfeld for the spinach *rbc*L gene, Drs J D Palmer and W F Thompson for the spinach plastid clone bank and Dr A Neale for his tobacco chitinase clones.

I am indebted to my family and friends for their support and encouragement.

Abbreviations

bp: base pairs

cDNA: DNA complementary to RNA

dH_2O: deionized water

DNA: deoxyribonucleic acid

EDTA: (ethylenedinitrilo)tetraacetic acid

g: gram

IPTG: isopropyl-B-D-thiogalactopyranoside

kbp: kilobase pairs

M: molar

mg: milligram

ml: milliliter

mRNA: messenger RNA

O.D.: optical density

ORF: translational open reading frame.

PCR: polymerase chain reaction

r.p.m.: revolutions per minute

RNA: ribonucleic acid

rRNA: ribosomal ribonucleic acid

SDS: sodium dodecyl sulfate

Tris: 2-amino-2-(hydroxymethyl)-1,3-propanediol

µCi: microcurie

µg: microgram

µl: micro litre

v/v: volume per volume

v/w: weight per volume

X-gal: 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside

Chloroplast gene nonmenclature is as recommended by Hallick and Bottomley (1983).

Publications

Ayliffe M.A., Timmis J.N., Steele Scott N. (1988). Homologies to chloroplast DNA in the nuclear DNA of a number of Chenopod species. *Theor. Appl. Genet.* 75: 282-285.

Ayliffe M.A., Timmis J.N. (1992). Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. *Theor. Appl. Genet.* - in press.

Steele Scott N., Timmis J.N., Ayliffe M.A. (1991). Chloroplast genes in higher plants. *Aust. J. Plant Physiol.* 18: 533-541.

Timmis J.N., Ayliffe M.A. (1991). Extranuclear DNA and its use in systematics. *Aust. J. Syst. Bot.* 3: 137-144.

CHAPTER 1

Introduction

The eucaryotic cell is divided into a variety of subcellular structures distinct from the cytoplasm termed organelles. Each organelle possess a specific function in cellular metabolism. Of the array of cellular organelles only three contain genetic information in the form of deoxyribonucleic acid (DNA); these being the nucleus, the mitochondrion and the plastid, although it has been proposed that basal body filaments of eucaryotic flagella and cilia may also contain DNA (Hall *et al*, 1989; Goodenough *et al*, 1989).

1.1: **The Nucleus**

Under the light microscope the most conspicuous organelle within the eucaryotic cell is the nucleus. The nucleus contains the vast majority of DNA and ultimately controls all aspects of cellular function and biogenesis. Within the nucleus DNA macromolecules are complexed with proteins to form the chromosome. The amount of DNA in the nucleus and the presence or absence of the nucleus itself, is dependant upon the stage of the cell cycle. Associated with the chromosomes is a suborganelle termed the nucleolus, which results from the processing of ribosomal ribonucleic acid (rRNA). Surrounding the nucleus are two membranes which fuse to form nuclear pores that regulate the movement of material between the cytoplasm and nucleus.

1.2: **The Mitochondrion**

The mitochondrion is an almost ubiquitous organelle amongst eucaryotic organisms and is responsible for cellular energy production *via* oxidation of tricarboxyclic acids (ie. the Krebs cycle). Specifically pyruvate generated by glycolysis is completely oxidized to form carbon dioxide. This reaction is coupled to the reduction

of molecular oxygen to form water with the net production of chemical energy in the form of adenosine triphosphate.

The mitochondrion is bound by two membranes with the inner membrane forming complex invaginations called cristae. Residing in the mitochondrial matrix is the mitochondrial DNA and all the necessary transcription and translation apparatus required for expression of mitochondrial encoded genes. Each mitochondrion contains multiple copies of the mitochondrial genome and most cells contain multiple mitochondria.

Although the mitochondrion is common to most eucaryotes, the mitochondrial genome is highly heterogeneous showing little conservation between different taxa. This genomic variation is reflected predominantly in size and structural variation rather than extreme differences in coding content, although some variation in gene content is evident (for review see Gray, 1989a).

Animal mitochondrial genomes are amongst the smallest and most conserved, representing an extreme example of genetic economy. The sequenced human mitochondrial genome consists of 16569 bp in which are encoded 13 protein genes, 22 tRNA molecules and two rRNA genes. Genes encoded in the human mitochondrial genome contain no introns. A pair of overlapping genes are also present (Fearnely and Walker, 1986).

Rearrangements in animal mitochondria are rare as would be expected from the high density coding capacity of the genome. However animal mitochondrial genomes often show very rapid rates of DNA sequence evolution, although different genes within the genome diverge at different rates. For example primate mitochondrial tRNA genes are thought to diverge 100 times faster than their nuclear encoded counterparts, even though these mitochondrial tRNA genes are diverging at only half the rate of some mitochondrial protein genes. This rapid sequence change is attributable to a high proportion of silent base substitutions, a high nucleotide transition to transversion ratio (10 to 1) and a strong bias for C and T interchange in the L strand (Gray, 1989a).

Fungal mitochondrial genomes, unlike their animal counterparts, show a large

variation in size (17.6 kbp to 115 kbp)(Gray *et al*, 1989a). This size variation is due predominantly to the amount of noncoding DNA contained in each genome; including the presence or absence of introns, variation in intergenic spacer lengths, the number of repeated sequences present and to sequence duplications.

As many fungi are capable of anaerobic metabolism, some fungal mitochondrial genomes can undergo extensive rearrangement as a functional mitochondrion is not essential. The most studied examples of this phenomenon occur in yeast where loss of wild type mitochondrial sequences can reduce the entire genome to a subgenome of only 35 bp (Fangman *et al*, 1989). The excision of this wild type mitochondrial DNA is *via* homologous recombination between intergenic spacers. Yeast mitochondrial DNA intergenic spacers are very A and T rich and are punctuated by short regions (10 to 60 bp) of G+C rich sequence. These short G+C rich elements appear to act as hotspots for intragenic homologous recombination resulting in loss of portions of the genome (Dieckmann *et al*, 1987). Rearrangement is further facilitated by the circularity of the mitochondrial genome and a high rate of recombination.

The largest and most complex of any mitochondrial genomes occur in higher plants, which show a size range from 208 kb in *Brassica hirta* (Palmer and Herbon, 1987) to 2400 kb in *Cucumis melo* (Ward *et al*, 1981). Plant mitochondrial genomes often show a lot of physical heterogeneity and the genome is thought to exist as a circular "master" chromosome which is frequently converted to smaller subgenomic circular and linear molecules, by recombination between repeated sequences (Lonsdale, 1984; Palmer, 1984). The stoichiometry of these subgenomic complexes is often quite complex (Small, 1989).

Plant mitochondrial genomes are estimated to encode approximately 20 polypeptide products in addition to tRNA's and rRNA's (Levings and Brown, 1989; Breiman and Galun, 1990) and their actual coding content is very similar to that of the much smaller animal mitochondrial genomes. The bulk of the genome is therefore noncoding DNA, including introns and repeated DNA sequences, with usually solitary genes being scattered throughout the genome. To date gene maps for different plant

mitochondrial genomes are very different (Palmer and Herbon, 1987; Gray, 1989a).

Unlike animal mitochondrial genomes, plant genomes evolve rapidly in structure but show slow rates of sequence divergence. Palmer and Herbon (1988) examined the mitochondrial genomes from six species of crucifers and found that each genome was structurally different but all had very similar primary sequence. Sequence divergence of 1 - 8 changes per 1000 nucleotide pairs was observed between the mitochondrial genomes of different species, which represents a point mutation rate 100 times slower than that of animal mitochondrial DNA. This slow rate of sequence divergence was evident in both coding and noncoding genomic locations.

The most poorly studied group in terms of mitochondrial genome structure and diversity are the protists (Gray, 1989a). Protozoan mitochondrial genomes are often linear with a size range of 15.8 kbp in *Chlamydomonas renhardtii* (Boer *et al*, 1985) up to approximately 60 kbp.

In addition to the complex variation amongst mitochondrial genomes, often associated with the mitochondria are extra autonomous nucleic acids, usually termed plasmids. These plasmids are found most commonly in plant and fungal mitochondria and consist of either DNA (Meinhardt *et al*, 1990; Kemble *et al*, 1980; Shikanai *et al*, 1987) or double or single stranded RNA (Levings and Brown, 1989; Kims *et al*, 1989). The size range of these molecules is quite large with some reported in excess of 10 kbp (Meinhardt *et al*, 1990) and structurally these molecules can be either circular or linear. No clear biological function has been attributed to these plasmids but they have been implicated in a variety of processes including plant cytoplasmic male sterility, senescence in *Neurospora* and in the basis of the killer phenotype in yeast (Meinhardt *et al*, 1990). The evolutionary origin of these molecules is also unknown and they have been compared with viruses, retrotransposons and introns (Kupier *et al*, 1988) and could possibly represent some form of horiziontal gene transfer to the mitochondrion (Gray *et al*, 1989b).

1.3: **The Plastid**

Photosynthetic eucaryotes are distinguished from heterotrophic eucaryotes by possessing an additional organelle the plastid. The term plastid encompasses a variety of forms of the same organelle, with each plastid type possessing a different function. For example the chloroplast is an actively photosynthetic form of plastid, the amyloplast is a plastid involved in starch synthesis and storage and the chromoplast is a carotenoid containing plastid responsible for some fruit and flower colours. These different plastid types are considered the same organelle as each is derived from the same progenitor, the proplastid. In addition plastids are generally interconvertible; that is some plastid types can be shown to derive from or form another plastid type. In general all the plastids in a single plant cell are of the same type and the form taken by the plastids of a cell is determined by the cell type. All the genetic material utilized in this study was obtained from green leaf tissue of which the chloroplast is by far the most abundant plastid.

The chloroplast is the site of photosynthesis, where light energy is utilized in the production of organic carbon from carbon dioxide. The chloroplast consists of a double membrane envelope surrounding a fluid filled cavity (the stroma) in which reside vesicles of photosynthetic membranes called thylakoids. Within the thylakoid membrane are photosynthetic pigments (chlorophylls) which absorb light energy and drive intermolecular transfer of electrons to produce ATP and NADPH (the so called light reactions of photosynthesis). This chemical energy is subsequently utilized in the synthesis of carbohydrates from carbon dioxide (the dark reactions of photosynthesis) within the chloroplast stroma.

Residing within the chloroplast stroma is the chloroplast genome and all the necessary transcription and translation apparatus for expression of genes encoded by this DNA molecule. Like the mitochondrion each plastid possesses multiple copies of the plastid genome. These numerous plastid genomes are not distributed uniformly throughout the stroma, but are organized into multicopy groups called nucleoids.

As is the case with the mitochondrion each plant cell possesses multiple chloroplasts and each chloroplast possesses multiple chloroplast genomes. The number of chloroplasts per cell and the number of genomes per chloroplast is dependant upon the age of the tissue. In spinach, mature leaf cells contain up to 200 chloroplasts per cell and 5000 copies of the chloroplast genome while very young leaf cells contain approximately 10 chloroplasts each containing 70-100 copies of the plastid genome. (Scott and Possingham, 1980).

Unlike the mitochondrial genome the plastid genome is highly conserved in size and structure. The plastid genomes of most land plants and green algae are circular molecules with a size range of 120-217 kbp and consist of two large inverted repeats separated by a large and small single copy region (Palmer, 1985b). Exceptions to this rule include a legume group in the family Fabaceae (Palmer and Thompson, 1982) which lack the inverted repeat and the plastid genomes of some brown algae which are bimolecular (de Goer *et al*, 1988; Markowicz *et al*, 1988). Recently pulse-field gel electrophoresis has suggested that plastid DNA may also exist in linear multimeric forms (Deng *et al*, 1989; Bendich *et al*, 1990).

To date the plastid genomes of three plant species have been sequenced in their entirety. These being the 155,844 bp plastid genome of tobacco (Shinozaki *et al*, 1986), the 134,525 bp rice plastid genome (Hiratsuka *et al*, 1989) and the 121,024 bp plastid genome of the liverwort *Marchantia polymorpha* (Ohyama *et al*, 1988a). Comparisons between the three genomes show that they are very similar in gene content, gene order and gene sequence (Sugiura, 1989). Size difference between these genomes is largely accounted for by differences in the size of the inverted repeat and intergenic spacer regions. Little sequence homology exists between the intergenic spacer regions of rice and tobacco when compared with the liverwort *Marchantia polymorpha* (Shimada and Sugiura, 1991).

Although plastid genomes of different plant species are apparently very similar in gene content, the organization of genes upon the circular chromosome does show variability (Palmer, 1985 and 1985b). The presence of short repeated sequences in some

plastid genomes suggests that genomic rearrangements may occur by duplication, deletion and recombination (Steele Scott *et al*, 1991).

The plastid genome of land plants encode four different rRNA genes (localized on the inverted repeat), 30 different tRNA molecules and over 62 polypeptide genes (including putative genes) (Shimada *et al*, 1991; Sugiura, 1989) Included amongst these polypeptide encoding genes are genes for ribosomal proteins, RNA polymerase subunits, photosystem components and polypeptides with homology to the mitochondrial NADH dehydrogenase complex. However many open reading frames still remain to be characterised and the potential number of genes encoded in plastid DNA is estimated between 120-125 as identified by translational ORF (Ohyama *et al*, 1988b).

Genomes of interrelated plastid types are considered identical in sequence and the plastid genomes within an individual plant are usually also considered homogeneous (Steele Scott *et al*, 1991).

1.4: **Nuclear Dependence Of Cytoplasmic Organelles**

Although the plastid and mitochondrion are "semi-autonomous" in that they are capable of replication independent from that of the nucleus, both organelles are reliant upon nuclear expression for function and biogenesis. This nuclear dependence is due to the nucleus encoding the vast majority of the genes that organelles require for effective function and biogenesis (Breiman and Galun, 1990). A large proportion (95%) of mitochondrial polypeptides are encoded within the nuclear genome by approximately 400 genes (Lestienne, 1989), in addition to tRNA molecules in some organisms (Gray, 1989a; Marachal-Drouard *et al*, 1990), while 80-90 % of plastid proteins are nucleus encoded (Palmer, 1990).

These nucleus encoded mitochondrial and plastid genes are transcribed within the nucleus, translated in the cytoplasm and the polypeptide or RNA product imported into the relevant organelle. Protein import into both the mitochondrion and plastid is generally facilitated by the presence of a transit peptide sequence on the NH_2 terminus

of the precursor polypeptide, which targets the molecule to a particular location in the specific organelle. Upon entry into the organelle this transit peptide sequence is cleaved to form the mature peptide (Flugge, 1990).

Frequently cytoplasmic organellar proteins consist of multimeric complexes of both nuclear and organelle encoded polypeptides. After entry into the relevant organelle the nuclear encoded polypeptides are complexed with native organellar peptides, occasionally with the assistance of an additional nuclear protein termed a "chaperonin" (Hemmingsen *et al*, 1988; Ellis, 1987) to assemble the final organelle protein.

Therefore very complex gene regulatory processes exist to maintain the functional relationships between the nucleus and these two cytoplasmic organelles (Nagley, 1991; Burgess and Taylor, 1988; Gruissen, 1989).

1.5: **Origin of the mitochondrion and plastid: the Endosymbiont Theory**

The most widely accepted theory for the evolutionary origin of the mitochondrion and plastid is the endosymbiont theory (or serial endosymbiont theory). This proposes that both the mitochondrion and the plastid evolved from an ancient symbiotic relationship between procaryotic organisms which together constituted the primitive eucaryotic cell. Specifically the plastid is thought to have evolved from a cyanobacterium and prior to this event the mitochondrion from a purple photosynthetic bacterium (Gray, 1989b). Subsequently during endosymbiosis these procaryotic organisms evolved to become an integral part of the host cell in the form of organelles. Given that eucaryotic cells are now completely reliant upon nuclear expression for the vast majority of their protein (and some RNA) products, a massive transfer of genetic information from these procaryote cells to the host cell nucleus most probably accompanied endosymbiosis. Presumably this transfer occurred in the early stage of endosymbiosis as all eucaryotic plastid and mitochondrial genomes possess very similar gene contents (Palmer, 1990).

A symbiotic origin for organelles was proposed as early as 1883 (Schimper), but a formal proposal of the endosymbiont hypothesis was put forward by Margulis in 1970

(Taylor, 1988). Subsequently a large amount of ecological, morphological, biochemical and genetical evidence has been collected which supports the endosymbiont theory (Weeden, 1981; Delihas and Fox, 1988; Gray, 1989b; Tanaka *et al*, 1986; Shih *et al*, 1986; Jolley *et al*, 1980; Schenk *et al*, 1988; Taylor, 1988). Although this hypothesis is now essentially biological dogma, controversy still arises as to whether either the plastid or mitochondrion result from a monophyletic or polyphyletic evolutionary event. A polyphyletic origin of the mitochondrion is suggested by the structural diversity and rRNA sequence comparisons between animal and plant mitochondrial genomes. In contrast the similar gene content of all mitochondrial genomes suggests a monophyletic origin (Gray, 1989a). Similarly the existence of three different photosynthetic pigments in plant and algal plastid has been suggested to be due to a polyphyletic plastid origin (Gray, 1989b; Turner *et al*, 1989; Morden *et al*, 1989; Penny, 1989; Taylor, 1988). For organisms whose plastids are surrounded by more than two membranes it has been proposed that the initial endosymbiont precursor of these plastids was a eucaryotic algae rather than a photosynthetic procaryote (Douglas *et al*, 1991; Penny and O'Kelly, 1991).

Therefore underlying the endosymbiont theory is the assumption that genetic information from both the plastid and mitochondrion procaryotic progenitors has been transferred to the nuclear genome. This genetic information has then become an integral part of the nuclear genome and subsequently evolved to form nuclear-encoded plastid and mitochondrial genes. The net result was the formation of modern eucaryotic cytoplasmic organelles which are heavily dependant upon the nucleus for biogenesis and function.

1.6: **Transfer of Genetic Information Between Organelles**

Two lines of evidence demonstrate that genetic information has been, or still is being, transferred between organelle genomes. The first is the observation that some plastid and mitochondrial genes are nucleus encoded in some species, but encoded within the cytoplasmic organelle genomes of others. The second line of evidence is the

occurrence of DNA sequences common to more than one organelle genome, or so called promiscuous DNA sequences which are the subject of this thesis.

Several examples of genes with different organelle locations have been observed. In yeast the gene encoding the mitochondrial ATP synthase subunit 9 protein (*atp*9) is located in the mitochondrial genome, whereas in humans, mouse and the fungus *Podospora anserina* this gene is located in the nuclear genome (Brown *et al*, 1984; Ridder *et al* , 1991). Thus during evolution, *atp*9 presumably has been transferred from the mitochondrial genome to the nucleus in some species and lost from the mitochondrial genome upon expression of the nuclear gene. In the mitochondrial genomes of both *Neurospora crassa* (van den Boogaart *et al*, 1982) and *Aspergillus nidulans* (Brown *et al*, 1984) a copy of the *atp*9 gene still exists, inspite of the fact that this polypeptide is coded by a nuclear gene in both these species. The later fungal species may represent an intermediary state of this gene transfer process, although the high amino acid sequence homology shared by these mitochondrial genes suggest they may still be biologically active (Brown *et al*, 1984).

Within leguminous plants two examples of recent gene transfer from both the plastid and mitochondrial genome to the nucleus have been reported. The plastid ribosomal protein gene CL22 (*rpl*22) is present in the plastid genomes of all plants examined except legumes, and a functional copy of this gene is present within the pea nuclear genome (Gantt *et al*, 1991). Similarly the mitochondrial *cox*II gene has transferred from the mitochondrion to the nucleus during the evolution of flowering plants between 60-200 million years ago (Nugent and Palmer, 1991). Most legumes have a copy of the *cox*II gene in both organelle genomes, but the expression of these genes is not simultaneous. As the nuclear version of the *cox*II gene more closely resembles the edited *cox*II mRNA of the mitochondria it would appear that transfer of this gene to the nucleus has been *via* a processed RNA intermediate. Since transfer of the *cox*II and *rpl*22 gene to the nuclear genome both genes have gained an additional exon encoding a putative transit peptide, which is separated from the mature protein coding region by an intron.

Several other examples of evolutionary transfer of plastid genes to the nuclear genome have been identified. The plastid protein synthesis elongation factor Tu (*tuf*A) is located within the plastid genome of green algae, but in the nuclear genome of land plants (Bauldauf and Palmer, 1990). This plastid gene is believed to have transferred to the nucleus early in the evolution of the Charophyceae, but with chloroplast copies of this gene still being retained in some of the subsequent diverging lineages (Baldauf *et al*, 1990). The small subunit of the enzyme ribulose-1,3-bisphosphate carboxylase and the plastid ferredoxin gene are located on the cyanelle genome of the protozoan *Cyanophora paradoxa*, but in the nuclear genome of all land plants and green algae examined (Heinhorst and Shively, 1983; Bayer and Schenk, 1989). Associated with the plastid encoded *rbc*S gene of the chromophytic marine diatom *Cylindrotheca* is a short open reading frame encoding an acyl carrier protein that is highly homologous to the nuclear encoded acyl carrier proteins from plants (Hwang and Tabita, 1991). Similarly the *rpl*35 gene has been identified on the cyanelle genome of *Cyanophora paradoxa*, but is apparently encoded in the nuclear genome of spinach (Smooker *et al*, 1990a). Located in the plastid genome of *Acetabularia mediterranea* is a DNA sequence homologous to the *Drosophilia* *per* locus. In contrast higher plants show homology to this sequence only in the nuclear genome (Li-Weber *et al*, 1989).The plastid glyceraldehyde-3-phosphate dehydrogenase is composed of two subunits encoded in the plant nuclear genome. As the amino acid sequence of these two polypeptides is more closely related to that of thermophilic eubacteria than to an analogous cytoplasmic glyceraldehyde-3- phosphate dehydrogenase, these nuclear genes may be the result of an ancient endosymbiotic gene transfer event(s) (Martin and Cerff, 1986; Brinkman *et al*, 1987; Liaud *et al*, 1990; Shih *et al* , 1986). Less clear in origin is the plastid ribosmal protein L21 gene (*rpl*21) which is localized on the plastid genome of *Marchantia polymorpha* and in the nuclear genomes of higher plants, as it has been suggested that this nuclear gene is of mitochondrial rather than plastid origin (Martin *et al*, 1990; Smooker *et al*, 1990b).

The variation in gene content of both the plastid (Shimada and Sugiura, 1991) and mitochondrial genomes (Nugent and Palmer, 1991) of different species therefore suggests gene transfer from organellar genomes to the nucleus. However some variability in gene content could be the result of either complete loss of a gene within a species or the acquisition of a nuclear gene in the organellar genome, rather than represent the evolution of a nuclear encoded homologue. Alternatively nuclear encoded organelle genes may have evolved from duplicated pre-existing nuclear genes.

1.7: **Promiscuous DNA Sequences: Plastid Sequences Present Within The Mitochondrial Genome**

Present within the mitochondrial genome of many (if not all) higher plants is a range of DNA sequences with significant homology to DNA sequences that are actively transcribed and translated within the plastid genome. These "plastid-like" DNA sequences found in the mitochondrial genome are thought to be the result of transfer of DNA from the plastid genome and subsequent integration into the mitochondrial genome. Sequences common to more than one organelle genome have been named "promiscuous DNA" sequences (Ellis, 1982).

A 12 kbp DNA sequence present within the maize mitochondrial genome was the first and has remained the largest reported of these promiscuous DNA sequences (Stern and Lonsdale, 1982). This fragment was shown by hybridization and restriction mapping to be homologous to a region of the maize plastid genome encoding a variety of genes including the 16s rRNA and two tRNA molecules. Subsequently DNA hybridization studies identified plastid DNA homologies in the mitochondrial genomes of a wide variety of plant species including corn, mungbean, spinach, pea, watermelon, squash, cucumber, muskmelon and *Chenopodium album* (Stern and Palmer, 1984; Lonsdale *et al*, 1983; Bettini *et al*, 1988; Whisson and Steele Scott, 1985). Therefore these promiscuous DNA sequences are a phenomenon generally observed in mitochondrial genomes throughout the higher plant kingdom and involve diverse regions of the plastid genome.

More detailed hybridization studies have shown that these plastid DNA sequences are dispersed throughout the plant mitochondrial genome and closely related species may possess different complements of these sequences. Within the 327 kbp spinach mitochondrial genome at least 12 separate regions of the genome have been identified as having homology to plastid DNA (Stern and Palmer, 1986), while eight separate plastid DNA homologies are dispersed throughout the sunflower (*Helianthus annuus*) mitochondrial genome (Siculella and Palmer, 1988). A similar dispersed pattern of plastid DNA sequence homologies has also been identified in the mitochondrial genome of crucifer species (Nugent and Palmer, 1988). Examination of the mitochondrial genomes of six crucifer species revealed that four closely related *Brassica* species all possessed the same 12-14 kbp of plastid DNA dispersed throughout the same 11 mitochondrial genomic locations. However one of these species possessed an additional major region of plastid homology within its mitochondrial genome. Conversely two more distantly related crucifer species contained only 5-7 kbp of plastid DNA within their mitochondrial genomes. It was concluded that 2-6 % of the crucifer mitochondrial genome is of plastid origin, and that the plastid DNA sequences common to all six species represent a sequence transfer event that took place over 30 million years ago (Nugent and Palmer, 1988).

Although different plant mitochondrial genomes contain different amounts of plastid derived sequences there does not appear to be a correlation between genome size and amount of integrated plastid DNA (Pring and Lonsdale, 1985).

Several of these mitochondrial DNA sequences with plastid DNA homology have been cloned and sequenced in their entirety. Moon *et al* (1988) identified a 6.9 kbp fragment within the rice mitochondrial genome that encodes for a cluster of plastid genes; these being *rbc*L, *atp*B, *atp*E, *trn*M, *trn*V and *rpl*2. However within the rice plastid genome the *rpl*2 gene is not located near the other plastid genes identified within this mitochondrial sequence. As both the *rpl*2 gene and *rbc*L gene are flanked by 151 bp repeats in the plastid genome , the authors propose that this rearranged plastid gene cluster present in the mitochondrial genome is the result of homologous recombination

between direct repeats. Whether this recombination event occurred during or after transposition of these plastid genes to the mitochondrial genome is unknown.

In the mitochondrial genome of *Oenothera*, Schuster and Brennicke (1987a) identified a 2081 nucleotide fragment showing 94 % sequence homology to the plastid 4.5s rRNA gene, the 3' half of the 23s rRNA gene and the intergenic region between these genes. This sequence is believed to have been involved in an intramitochondrial rearrangement with the gene encoding for the putative mitochondrial small ribosomal protein S13. Also identified within the mitochondrial genome of *Oenothera* is a region that contains 758 nucleotides of high sequence homology to a portion of the plastid *rps*4 gene, the entire plastid $tRNA^{ser}$ gene and the intergenic region separating these two plastid genes. Seven hundred and sixty two nucleotides beyond the border of plastid homology lie 528 nucleotides showing 91% homology with the maize nuclear encoded 18s rRNA sequence and separating these plastid and nuclear sequences is an ORF encoding a polypeptide of 142 amino acids with significant homology to a reverse transcriptase (Schuster and Brennicke, 1987b).

Similarly an internal part of the plastid *atp*A gene has been identified in the wheat (*Triticum aestivum*) mitochondrial genome (Jubier *et al*, 1990). This 1106 bp insert showed 96% homology to the corresponding wheat plastid *atp*A gene differing only by 29 nucleotide changes and two small deletions of 6 bp and 4 bp from the plastid sequence. Bordering this plastid insert were two 9 bp direct repeats [TAGC(A/T)TCAA]. This plastid DNA insert is present in the mitochondrial genomes of several other *Triticum* species, but not within the maize mitochondrial genome. Therefore this transfer of DNA from the plastid genome to the mitochondrial genome probably occurred subsequent to the divergence of wheat and maize, but prior to the appearence of a common ancestor of *Triticum* species.

In addition to the 12 kbp plastid DNA fragment located within the maize mitochondrial genome (Stern and Lonsdale, 1982), a 1270 bp fragment encoding the 3' end of the plastid 23s rRNA gene, all of the plastid 4.5s and 5s rRNA genes and an almost complete copy of the plastid tRNA-arg gene, has been identified (Braun and

Levings, 1986). When compared with the appropriate region of the maize plastid genome, this mitochondrial fragment shows 99.2% sequence homology (Fejes *et al*, 1988).

Episomal nucleic acids of unknown origin within the plant mitochondria also have regions of DNA sequence homology to the plastid genome, although these episomes have never been identified within higher plant plastids. Present within the maize mitochondrial linear S1 plasmid is a segment of 420 nucleotides homologous to the 3' end of the plastid *psb*A gene (Sederoff *et al*, 1986). Scattered amongst this plastid sequence are several 9-12 bp insertions that appear to have arisen by a duplication of immediate surrounding sequences and are therefore reminiscent of plant transposable element footprints. However as this region of plastid DNA homology was not compared with the maize plastid genome the exact nature of these insertions could not be defined. A linear plasmid found in the mitochondria of *Brassica* has been shown to possess homology to the plastid genome by hybridization studies (Turpen *et al*, 1987).

Plant mitochondrial genomes encode numerous tRNA genes, all presumably essential for translation of mitochondrial polypeptide genes. However these tRNA genes appear to consist of two distinct populations; those showing 65-75% homology to plastid tRNA genes and those showing 95% and greater homology to their plastid counterparts. As the intergenic regions of this later population of tRNA genes often also show significant sequence homology to the plastid genome and a conserved gene order, these mitochondrial tRNA genes are believed to be of plastid origin. These genes represent one of the few examples of expression of a promiscuous DNA sequence.

The first example of a transcribed promiscuous plastid tRNA gene was a $tRNA^{trp}$ gene localized in the wheat mitochondrial genome (Marechal *et al*, 1987). This mitochondrial gene showed 97% homology with its plastid counterpart and 136 bp upstream of this gene were 23 nucleotides that were completely homologous to the 3' end of the plastid $tRNA^{pro}$ gene. Within the plastid genome of all higher plants analysed the $tRNA^{pro}$ and $tRNA^{trp}$ genes occur in tandem. Therefore it was concluded that this mitochondrial encoded $tRNA^{trp}$ gene is of plastid origin but it is present and

transcribed in wheat, bean and maize mitochondria. Within maize mitochondria this gene arrangement is identical to that observed in wheat, except that the $tRNA^{trp}$ gene is located on a 2.3 kbp linear plasmid.

Also within the maize mitochondrial genome a $tRNA^{cys}$ gene is transcribed that shows 99% sequence homology to the wheat plastid analogue (Wintz *et al*, 1988b). Upstream of the gene very little sequence homology with the wheat plastid genome is observed, whereas 73% homology to the plastid genome is present downstream. Subsequently it has been shown that the maize mitochondrial genome encodes six functional tRNA genes of plastid origin (Sangare *et al*, 1990).

Extensive analysis of tRNA molecules found in the potato mitochondrion suggest that five out of 31 identified tRNA molecules are transcribed from promiscuous plastid genes inserted into the mitochondrial genome (Marechal *et al*, 1990). "Plastid-like" tRNA molecules or tRNA genes have been described in various plant mitochondria specific for asparagine, cysteine, histidine, methionine (elongator), phenylalanine, serine (GGA) and tryptophan (Iams *et al*, 1985; Wintz *et al*, 1988a, 1988b; Joyce and Gray, 1989; Sangare *et al*, 1990). However different plant species apparently possess different complements of tRNA molecules encoded by genes of plastid origin (Marechal-Drouard *et al*, 1990)

As may be expected the plant mitochondrial genome also contains plastid derived sequences that encode for tRNA genes that are either nonfunctional or truncated within the mitochondria (Dron *et al*, 1985; Moon *et al*, 1988; Stern and Lonsdale, 1982; Sangare *et al*, 1990).

1.8: **Nuclear DNA Sequence Homologies In The Mitochondrial Genome**

Very few nuclear DNA sequences have been identified within mitochondrial genomes. Some of the smaller animal and insects mitochondrial genomes have been sequenced in their entirety and lack promiscuous sequences of either plastid or nuclear origin. However no systematic search of the larger plant mitochondrial genomes for DNA sequences of nuclear origin has been undertaken. Such a systematic search will be

complicated by the presence of mitochondrial and plastid sequences in the nuclear genome and the presence of plastid sequences in the mitochondrial genome. It has been predicted that much of the plant mitochondrial genome will eventually be shown to be of nuclear origin (Palmer, 1990).

The only clear example of a nuclear DNA sequence within a mitochondrial genome is the identification of 620 nucleotides of the nuclear 18s rRNA gene within the mitochondrial genome of *Oenothera* (Schuster and Brennicke, 1987b). Incorporated into this sequence is an expansion segment of 160 nucleotides which is unique to nuclear rRNAs. This sequence lies adjacent to an ORF encoding for a putative reverse transcriptase that is also postulated to be of nuclear origin (Schuster and Brennicke, 1987b). Beyond this reverse transcriptase-like ORF is a region of homology to the plastid *rps*4 and tRNA^{ser} genes.

Senescence of kalilo variants of *Neurospora intermedia* involves the integration of a 9 kbp transposable element (kalDNA) into the mitochondrial chromosome (Bertrand *et al*, 1985). As this element has no sequence homology to the mitochondrial genome it is presumed to be of nuclear origin.

If a DNA sequence is identified within two or more organelle genomes it is usually presumed to have originated from the genome in which it is known to be functional. However as the vast majority of the nuclear genome, particular in plants, has no obvious function it will be difficult to identify the origin of these sequences if they are identified as promiscuous. For example Bernatsky *et al* (1989) have identified a 56 bp DNA sequence in both the nuclear and mitochondrial genomes of *Nicotiana alata*. As this sequence has an indeterminate function in both genomes its genomic origin remains an enigma.

1.9: Mitochondrial DNA Sequence Homologies In The Nuclear Genome

In comparison with both the mitochondrial and plastid genomes the nuclear genome is enormous. Of the vast array of genes and sequences within this organelle some are of mitochondrial and plastid origin. Promiscuous mitochondrial DNA

sequences have been identified within the nuclear genomes of a wide range of organisms.

Within the yeast nuclear genome, Farrelly and Butow (1983) identified a 289 bp fragment of mitochondrial origin. 115 bp of this sequence showed greater than 85 % homology to the middle of a ribosomal associated protein gene (var1), while a further 123 bp is more than 80 % homologous to the 3' end of the mitochondrial cytochrome b apoprotein (cob/box) gene. Immediately adjacent to the 5' end of the nuclear homology to the cob/box gene was a short 266 bp stretch of homology to a portion of intron 5 of the cob/box gene, juxtaposed to a region homologous to the mitochondrial origin of replication. This arrangement does not represent a contiguous wild type mitochondrial DNA fragment and was proposed to be the result of integration of a petite mitochondrial genome into the yeast nucleus. That is, rearrangement of these mitochondrial genes took place prior to their transposition to the nucleus. In some yeast strains, closely linked to this sequence of mitochondrial origin is a tandem pair of transposable Ty elements.

Hybridization studies have further identified homologies between the yeast nuclear genome and at least eight portions of the yeast mitochondrial genome, including both exons and introns (Hudson *et al*,1985). Some of these homologies are also found in association with one or more Ty elements.

Unique evidence of sequence transfer from one organelle genome to another comes from the fungus *Podospora anserina*. This sequence transfer event is surprising in that it does not occur on an evolutionary time scale, but rather during the limited life-span of this organism. Associated with senescence of *Podospora* is the excision of a specific DNA segment from the mitochondrial genome (termed senDNA) and the subsequent amplification of these molecules (Vierny *et al*, 1982; Wright, 1982; Jamet-Vierny *et al*, 1980). At the final stage of senescence these plasmids constitute virtually all the DNA present within the senescent mitochondria and are also transposed and integrated into the nuclear genome (Wright and Cummings, 1983). In some cases this nuclear integration is as a highly repeated tandem array. However the observations of Wright and Cummings could not be reproduced by Koll (1986).

Within the locust, sea urchin and human nuclear genomes high sequence homology to mitochondrial rRNA genes has been observed. Isolated from the sea urchin nuclear genome (*Strongylocetrotus purpuratus*) was a sequence containing 342 bp of the 3' end of the mitochondrial 16s rRNA gene adjacent to the complete coding region of the cytochrome oxidase subunit I gene (*cox*I) (Jacobs *et al*, 1983; Jacobs and Grimer, 1986). Approximately 500 nucleotides upstream of this gene was a duplication of the *cox*I gene.

Similarly isolated from the locust nuclear genome was the 3' end of the mitochondrial large rRNA gene which showed 90 % homology to the true mitochondrial gene (Gellisen *et al*, 1983; Gellisen and Michaelis, 1988). It was estimated, from the number of positive clones in a genomic library, that several hundred copies of the mitochondrial rRNA gene occur within the *Locusta migratoria* nuclear genome.

Extensive analysis of mitochondrial DNA fragments inserted into the human nuclear genome has suggested that at least several hundred copies may be present within a haploid genome (Fukuda *et al*, 1985). These sequences involve various regions of the mitochondrial genome and appear to be dispersed throughout the nuclear genome, rather than representing a tandem array. Sequence analysis of four nuclear fragments with homology to a region of the mitochondrial genome encompassing unidentified ORF 4 and 5, and three tRNA genes suggests that three of these mitochondrial sequences were transferred to the nuclear genome approximately 40 million years ago. The remaining sequence is believed to have transferred to the nucleus subsequent to the divergence of chimpanzees and humans. Similarly two DNA fragments isolated from the human nuclear genome showed homology to the mitochondrial 16s rRNA gene and are believed to be the result of transposition events occurring 12 and 15 million years ago (Nomiyama *et al*, 1985; Tsuzuki *et al*, 1983). One of these sequences was a 0.4 kbp fragment showing 84 % homology to the 5' end of the 16s rRNA gene, while the other was a 1.6 kbp insert with 80 % homology to the entire mitochondrial 16s rRNA gene and the 3' end of the 12s rRNA gene.

After integration of these mitochondrial sequences into the nuclear genome, they are subject to the processes of nuclear genomic rearrangement. For example two mitochondrial-like fragments have been isolated from the human genome and both contained an insertion of a KpnI-like mobile genetic element. As these KpnI elements have not been identified in the mitochondrial genome, presumably insertion of these elements occurred after transposition of the sequences to the nucleus. One of these sequences which is believed to be the result of a transposition event from the mitochondria to the nucleus occurring over 40 million years ago, showed 73 % homology to the human mitochondrial 16s rRNA gene and URF2, and contained a 1.8 KpnI element integrated into this mitochondrial sequence homology (Nomiyama *et al*, 1984). The other fragment contained a 5.5 kbp KpnI element inserted into 7.5 kbp of nuclear DNA that showed 75 % homology to the mitochondrial URF4 and 5 and the *cox*I and *cox*II genes (Wakasugi *et al*, 1985).

A continual process of DNA transposition from the mitochondria to the nucleus has therefore occurred throughout evolution. Analysis of the sites of integration of these mitochondrial sequences in the human nuclear genome suggests that this process is essentially random and involves neither a specific nuclear target site nor any obvious recombinogenic sequences such as repeats.

Hybridization studies have also identified mitochondrial DNA sequence homologies within the nuclear genomes of a variety of organisms. Hadler *et al* (1983) demonstrated nuclear DNA sequence homologies to the mitochondrial D-loop and rRNA genes in rat, while Corral *et al* (1989) identified sequence homology within the rat nuclear genome to NADH dehydrogenase subunit 6 (*nd*6) and *cox*I-III. Similarly the maize mitochondrial episomal S1 plasmid has been shown to have homology to the maize nuclear genome (Kemble *et al*, 1983). However it should be noted that homology between the maize S1 plasmid and the maize plastid *psb*A gene has also been observed (Sederoff *et al*, 1986). Thus some of the homology between the S1 plasmid and the maize nuclear genome could be the result of homology between a promiscuous plastid

sequence in the nucleus and the promiscuous plastid sequence present in the S1 plasmid. Therefore care should be taken when interpreting hybridization data.

Several other episomal plasmid molecules found in mitochondria have been shown to hybridize to plant nuclear genomic DNA (Smith and Pring, 1987; Sakamoto *et al*, 1991; Crouzillat *et al*, 1989; Kanazawa *et al*, 1991). As some of these mitochondrial plasmids showed no homology to the mitochondrial genome, it has been postulated that these sequences may be of nuclear origin. Homology between these molecules and the nuclear genome could therefore be due to a nuclear origin of these episomes, rather than a transposition event from the mitochondria to the nuclear genome.

1.10: **Plastid DNA Sequence Homologies In The Nuclear Genome**

Plastid DNA sequence homologies were first identified within the spinach nuclear genome by DNA hybridization studies (Timmis and Steele Scott, 1983). These homologies were identified as several distinct nuclear DNA restriction fragments with homology to a single plastid DNA probe, thereby demonstrating multiple integration sites of a plastid sequence within the nuclear genome. Subsequently it was estimated that between 5-10 copies of potentially the entire plastid genome were integrated within a haploid spinach nuclear genome (Steele Scott and Timmis, 1984). These plastid sequence homologies appeared to be less than 2-3 kbp in length and generally located in heavily methylated regions of the genome. Some of the promiscuous plastid sequences found in the spinach nuclear genome are also present in the mitochondrial genome, demonstrating that certain sequences are common to all three genetic compartments (Whisson and Steele Scott, 1985).

The occurrence of plastid DNA sequence homologies within the plant nuclear genome was subsequently extended to five other members of the Chenopodiaceae, demonstrating that this is a widespread phenomenon (Ayliffe *et al*, 1988). However the number and types of nuclear homology were variable within a single plant family. In both *Chenopodium album* and *Atriplex cinerea* a nuclear EcoRI restriction fragment with plastid DNA sequence homology was identified that was identical in size to the

bona fide plastid EcoRI restriction fragment; suggesting the existence of large (in excess of 6 kbp) tracts of plastid DNA within these nuclear genomes.

Hybridization studies have further identified sequence homologies between the potato nuclear genome and at least a quarter of the potato plastid genome (du Jardin, 1990). These nuclear homologies were again identified in heavily methylated regions of the genome, and some were estimated to be in excess of 13 kbp. Substantial homology was also observed between the potato plastid and mitochondrial genomes.

Cloning and sequencing of the nuclear encoded plastid *Cab-7* gene of tomato (*Lycopersicon esculentum*) fortuitously identified a promiscuous plastid sequence within this gene (Pichersky and Tanksley, 1988). Residing in the third intron of this nuclear gene are two fragments that showed high sequence homology to the plastid encoded *psb*G gene. The first fragment of 133 bp was flanked by 11 bp direct repeats, while the later 107 bp fragment was flanked by 6 bp direct repeats. In both cases the repeat unit at the 3' end was a part of the plastid DNA insert, while the 5' repeat was of nuclear origin. These two sequences were greater than 80 % homologous and the smaller fragment was believed to have derived by a duplication of the larger. The initial transposition, from the plastid genome to the nuclear genome of this sequence, was estimated to have occurred subsequent to the divergence of the genus *Lycopersicon* from other genera of the *Solanaceae* but before radiation of species within the *Lycopersicon* genus.

Screening of two tomato genomic libraries covering 95% of the genome, with 58 kbp of DNA from the large single copy region of the plastid genome identified only a further two promiscuous plastid sequences within the nuclear genome (Pichersky *et al*, 1991). One of these sequences mapped to chromosome 9 and was a 202 bp fragment with high sequence homology to a region of the plastid *rps*18 gene. The other fragment of only 141 bp mapped to chromosome 5 and was homologous to a region of the plastid *pet*D gene. Both inserts were again flanked by short direct repeats.

As in tomato, the spinach nuclear genome contains short tracts of homology to the plastid genome. Three segments with high sequence homology to the *rbc*L gene

have been isolated from the spinach nuclear genome, that are 63, 150 and 214 bp in length (Cheung and Steele Scott, unpublished results). Flanking these sequences were short degenerate repeats ranging in size from 7-18 bp. Unlike the repeat motifs flanking the tomato promiscuous plastid sequences described by Pichersky *et al* (1991) none of these repeats was part of the plastid insert. Located further away from each of these sequences of plastid origin were juxtaposed short direct repeats reminiscent of the target site duplication caused by excision of transposable elements.

The spinach nuclear genome also contains a chimeric promiscuous plastid sequence consisting of three separate regions of the spinach plastid genome (Cheung and Steele Scott, 1989). This 3.4 kbp sequence showed greater than 99 % homology to the plastid *rps*2, *psb*D/C and *psa*A genes. Analysis of the plastid/plastid and nuclear/plastid junctions revealed no obvious structure elucidating the mechanism of transfer of this plastid sequence.

The rice nuclear genome contains a 1.4 kbp fragment with homology to the 3' end of the tobacco plastid *rps*12 gene and the entire *rps*7 gene (Kikuchi *et al*, 1987). This sequence was apparently highly variable in copy number in tissue culture and during plant development, ranging from 100,000 copies per haploid nuclear genome in rice embryo tissue to only 5-10 copies per genome in callus culture. This observation parallels two sequences present in both the mitochondrial and nuclear genomes of *Cucumis melo* that also showed variable copy numbers during callus formation (Grisvard *et al*, 1990).

1.11: **Promiscuous Sequences Present Within The Plastid Genome ?**

DNA sequences of either nuclear or mitochondrial origin have never been conclusively identified within a plant plastid genome; despite three plastid genomes having been sequenced in their entirety (Shinozaki *et al*, 1986; Hiratsuka *et al*, 1989; Ohyama *et al*, 1986). However some sequences have been suggested to be potentially of mitochondrial or nuclear origin.

The tobacco plastid genome contains six ORFs whose deduced amino acid sequences resembles components of the mitochondrial NADH dehydrogenase complex (*nd*1, 2, 3, 4, 4L, 5) (Shinozaki *et al*, 1986). Similar sequences have been identified within the plastid genomes of other plants (Sugiura, 1989) and such sequences may have originated in the mitochondria. However the widespread occurrence of these sequences in plant species, their sequence divergence compared with their mitochondrial counterparts and their known active expression in chloroplasts (Vera *et al*, 1990) argues against a mitochondrial origin. These two sets of organellar genes are definitely related, but they have long separate evolutionary histories, probably predating endosymbiosis (Nugent and Palmer, 1988; Schuster and Brennicke, 1988).

The origin of introns has long been debated with two main schools of thought. The "intron early" hypothesis suggests that introns were common early in evolution but were subsequently lost from eubacteria, remaining only in eucaryotes and a few archebacterial and bacteriophage genes. The "intron late" argument proposes that introns evolved after the divergence of eubacteria and eucaryotes and subsequently spread by lateral transfer to many different eucaryotic genes. This later hypothesis is supported by several examples of intron mobility (reviewed by Scazzocchio, 1989; Lambowitz, 1989; Rogers, 1989).

If the "intron late" hypothesis is correct plastid and mitochondrial genomes, which evolved from supposedly intron-less eubacteria, must have acquired introns presumably from the eucaryotic nucleus. Thus introns present in plastid encoded genes would actually represent promiscuous nuclear sequences. Recent evidence strongly contradicts this scenario, at least for group I introns.

Group I introns were discovered in the $tRNA^{leu}$ gene of eight cyanobacterial species and these introns are very similar in sequence and identical in location to the intron present in the plant plastid $tRNA^{leu}$ gene (Xu *et al*, 1990; Kuhsel, 1990). This finding strongly supports the "intron early" hypothesis suggesting that introns present in cytoplasmic organelles were acquired from their procaryote ancestors (Belfort, 1991).

Given that a massive transfer of genetic information from the procaryotic endosymbiont to the host cell nucleus occurred during the evolution of cytoplasmic organelles and the recent observation of eubacterial group I introns, Belfort (1991) has speculated that nuclear group I introns may have been acquired from cytoplasmic organellar genomes, although the author stressed that other potential routes of nuclear intron acquisition exist. Additionally group two introns which are found in both plastid and mitochondrial genomes have not yet been identified within the nuclear genome (Rogers, 1989).

More detailed examination of other plant and algal plastid genomes may also reveal sequences of potential nuclear or mitochondrial origin. The plastid genome of *Trifolium subterraneum* has been demonstrated as unusual when compared with other plant plastid genomes because it is a highly reorganized genome containing repeated and single copy DNA sequences which have not been identified in other plastid genomes (Milligan *et al*, 1989). These previously unidentified sequences are potentially of mitochondrial or nuclear origin. Given the extreme complexity of both plant nuclear and mitochondrial genomes and our limited knowledge of their sequence composition, the potential for some plastid sequences to be of promiscuous origin can not be discounted.

However promiscuous sequences have never been conclusively identified within a plastid genome and possibly do not exist in plastid genomes. The plastid genome is a highly conserved molecule and therefore very evolutionarily stable. A high proportion of the genome is either coding or functional so has limited potential sites for insertion of foreign DNA. This observation parallels the absence of promiscuous sequences in the extremely compact animal mitochondrial genomes. In addition the rapid loss of photosynthetic genes from the plastid genomes of parasitic plants (De Pamphilis and Palmer, 1990; Wimpee *et al*, 1991) suggests that the plastid genome is under strong selection for an overall reduction in genome size and is therefore not amenable for expansion by the introduction of foreign DNA sequences.

1.12: Mechanisms Of Promiscuous DNA Sequence Transfer

The mechanism(s) of transfer of genetic information from one organelle genome to another is essentially unknown and could involve more than a single mechanism. For example transfer of DNA sequences from either the plastid or mitochondria to the nucleus could be envisaged by degradation of either organelle and release of multiple organelle genomes into the cytoplasm (Stern and Palmer, 1984; Pring and Lonsdale; 1985). As an individual cell contains multiple plastids and mitochondria, the degradation of a single organelle would have little effect. The nuclear membrane is porous and appears to provide no significant barrier to transfer of a variety of RNA, DNA and protein molecules. Fragments or entire copies of these organelle genomes could integrate into the nucleus by either homologous or heterologous recombination. The observation of entire organelles within the nucleus supports this hypothesis (Brandes *et al*, 1965: Bloom, 1967).

The same process could obviously not account for the presence of nuclear and plastid sequences within the mitochondrial genome as the double membrane of the mitochondrial envelope would prove a significant barrier to passive uptake of either DNA or RNA. A number of reports have demonstrated that both RNA and protein-DNA complexes are imported into the mitochondrial matrix (Vestweber and Schatz, 1989: Hancock and Hajduk, 1990; Nagley, 1989; Suyama, 1986). Presumably these processes involve either some specific receptors or an active process analogous to bacterial transformation. Direct contact or fusion between the mitochondria and plastid may also facilitate nucleic acid transfer (Stern and Palmer, 1984; Pring and Lonsdale; 1985).

The multicopy status of the mitochondrial genome is significant in discussion of the fate of organelles with altered genomes. Presumably each promiscuous sequence represents a single insertion event within a single mitochondrial genome which then becomes fixed within the entire mitochondria population. There may be many transpositions only a few of which become fixed by fortuitously inserting into a mitochondrial genome destined for replication (Pring and Lonsdale, 1985; Lonsdale,

1985). Although in culture all the tobacco cell mitochondrial genomes apparently undergo replication (Infante and Weissbach, 1990).

The generally high sequence homology between promiscuous sequences and the genome from which they were derived suggests that, on an evolutionary scale, this sequence transfer between organelles is a frequent process. However only a single experiment has ever been able to estimate the frequency of transfer of a DNA molecule between organelles. An *in vivo* experiment in yeast demonstrates that transfer of a plasmid present within the mitochondria to the nucleus occurred at a frequency of 2×10^{-5} transfers per cell per generation (Thorsness and Fox, 1990). Conversely the reciprocal transfer from the nucleus to the mitochondria was at least 100,000 times less frequent. Significantly the nuclear genotype affected DNA transfer from the mitochondria to the nucleus by up to fourfold (Thorsness and Fox, 1990).

The movement of genetic material between organelles has been also suggested to occur *via* some type of vector sequence (Stern and Palmer, 1984: Pring and Lonsdale, 1985). Mitochondrial plasmids have been shown to have regions of DNA sequence homology to both plastid (Sederoff *et al*, 1986; Turpon *et al*, 1987) and nuclear genomes (Smith and Pring, 1987; Kemble *et al*, 1983). These molecules may be acting as vector sequences transferring genetic information from one organelle genome to another. The association of mitochondrial DNA inserts in the yeast nuclear genome with yeast transposable elements (Farrelly and Butow, 1983, Hudson *et al*, 1985; Pring and Lonsdale, 1985) also suggests some form of vector sequence. Viruses could transfer genetic information between organelles *via* a transduction-like process (Stern and Palmer, 1984). However of the numerous promiscuous sequences identified to date, few show any evidence of association with potential vector sequences.

Identification of 528 nucleotides of the nuclear small rRNA gene and the plastid *rps*4 and $tRNA^{ser}$ genes in the mitochondrial genome of *Oenothera*, separated by an ORF with homology to a reverse transcriptase, lead Schuster and Brennicke (1987b) to suggest that genetic information may be transferred between organelles *via* an RNA intermediate. Recent evidence firmly supporting this theory is the identification of the

mitochondrial *cox*II gene encoded within the legume nuclear genome (Nugent and Palmer, 1991). This nuclear gene more closely resembles the edited *cox*II mRNA than the mitochondrial encoded *cox*II gene, suggesting that it was derived from a processed RNA intermediate.

Upon reaching the organelle genome a promiscuous sequence has several potential methods of inserting into the genome. Homologous recombination could occur by fortuitous homology between the target site and integrating sequence eg. tRNA or rRNA genes. (Fejes *et al*, 1988; Curtis and Rawson, 1982). If transfer has been *via* an RNA intermediate, a process of reverse transcription must have preceded genomic integration (Schuster and Brennicke, 1987b, 1988; Nugent and Palmer, 1991). The mechanism involved can probably only ever be elucidated by analysis of the sites of integration of these sequences ie. the border region of the promiscuous sequence and the host genome. However junction analysis has two main limitations. Generally the original sequence of the target site into which the promiscuous sequence integrated is unknown, unless the target site can be unambiguously identified in a closely related species or strain lacking the insertion. Integration of foreign DNA into the nuclear genome can involve rearrangement of the target site in addition to up to several hundred base pairs of so called "filler DNA" being added (Gheysen *et al*, 1987; Sainsard-Chanet and Egel, 1990; Roth *et al*, 1989). Secondly, upon integration into either a mitochondrial or nuclear genome, a promiscuous sequence may be amplified and integrated into other regions of the host genome by a process unrelated to the initial transposition event. Therefore the border region of a promiscuous sequence may not represent the initial target site but rather be the result of a subsequent host genome duplication or rearrangement.

Analysis by Pichersky and Tanksley (1988) and Pichersky *et al* (1991) of the junction sites of four plastid DNA insertions within the tomato nuclear genome revealed a common motif in each insert. Flanking each insert are direct repeats ranging in size from 5-11 bp. Plant transposable element insertions result in the formation of direct repeats flanking the integrated element by a duplication of the target site of integration

(Flavell, 1988). However the repeat located at the 3' end of these plastid inserts is a contiguous part of the plastid sequence, while the repeat at the 5' end is not part of the plastid insert and is presumably nuclear in origin. Pichersky (1990) subsequently describes a model for insertion of these sequences involving homologous base pairing between a short single-stranded region of the inserting sequence and a short single-stranded region of the chromosome resulting from a staggered break. The other end of this staggered break is endfilled and ligated to the unpaired end of the inserting sequence.

1.13: **Expression Of Promiscuous Sequences**

After integrating into a new organelle genome a promiscuous sequence has several barriers preventing its expression. Firstly each organelle genome has its own unique transcription initiation systems. For example the promoter sequences of plastid genes are more closely related to procaryotic promoters than nuclear gene promoters (Kung and Lin, 1985; Hanely-Bowdoin *et al*, 1985). Even if transcription of a promiscuous sequence occurred it would still probably lack the necessary translation initiation sequences and therefore would not form a polypeptide. A mRNA transcribed from a promiscuous sequence may also be unstable within its new location. Futhermore as mitochondrial genes may not adhere to the universal genetic code and both mitochondrial and plastid transcripts require post-transcriptional editing by RNA processing the effective expression of "foreign" genes becomes even more unlikely (Covello and Gray, 1989; Gualberto *et al*, 1989; Hoch *et al*, 1991; Cattaneo, 1992). Coupled with these expression difficulties is the problem of sequence divergence. As promiscuous sequences are not essential for cell function fewer selective forces maintain their sequence integrity. Consequently the sequence of a promiscuous insert may rapidly drift, ultimately resulting in the destruction of any potential coding sequence. It has been estimated that such a sequence could exist for perhaps 10-20 million years before being rendered inactive by mutation (Gantt *et al*, 1991).

Inspite of these difficulties a few examples of either expression or potential expression of promiscuous sequences occur. The most definitive example is the identification of functional mitochondrial tRNA genes that are of plastid origin (Marechal-Drouard *et al*, 1990; Wintz *et al*, 1988; Joyce and Gray, 1989). In addition approximately 270 bp of the plastid *psb*A gene is thought to be expressed as part of a 800 bp mitochondrial transcript in atrazine resistant *Chenopodium album* plants (Bettini *et al*, 1988). Identified within the maize mitochondrial genome is a transcribed ORF that is formed from multiple intramitochondrial recombination events involving a promiscuous plastid sequence (Dewey *et al*, 1986). Transcription of the tomato *Cab-7* gene (Piechulla *et al*, 1991) would also involve the transcription of two small plastid sequences located in the third intron of this gene (Pichersky and Tanksley, 1988). However subsequent RNA processing would presumably result in the removal of these foreign sequences from the transcript. PCR analysis has shown that the *rpl*2 gene contained on the 12 kbp plastid DNA insert in the maize mitochondrial genome (Stern and Lonsdale, 1982) is also transcribed (Hoch *et al*, 1991).

More circumstantial is the apparent localization of the small subunit protein of the enzyme ribulose bisphosphate carboxylase within the mitochondria of *Ochromonas danica*; suggesting that this protein may be synthesized from a promiscuous sequence present within the mitochondrial genome (Lacoste-Royal and Gibbs, 1985). Homology between transcripts from the mitochondrial genome of *Brassica napus* and plastid DNA suggests the expression of promiscuous plastid sequences localized in the mitochondrial genome (Carlson *et al*, 1986). However plastid sequences in the mitochondrial genome of *Brassica campestris* do not appear to be transcribed (Makaroff and Palmer, 1987).

Of the variety of promiscuous sequences identified, very few show any evidence of expression. For a promiscuous sequence to evolve into a functional gene it must rapidly acquire the correct transcription and translation initiation sequences and often sequences to direct the product back to the specific organelle. Following the construction of this "new" gene its homologous counterpart, in the organelle genome

from which it was derived, must be inactivated to preserve the integrity of the new gene by selection.

1.14: **Summary**

Promiscuous sequences are sequences common to more than one organelle genome. They have been identified in the nuclear and mitochondrial genomes of a wide variety of species but never satisfactorily in plastid genomes or in the highly compact animal mitochondrial genomes. Sequences from all three genetic compartments have been identified as promiscuous. Promiscuous sequences are generally highly homologous to the organelle genome from which they were derived and vary in size from approximately 50 bp up to 12 kbp. The mechanism of transfer of these sequences is yet to be elucidated, although several potential mechanisms have been postulated. In a few rare instances some of the genes encoded by these sequences are capable of at least being transcribed within their new genomic location.

CHAPTER 2

2.1: Materials

General reagents

General laboratory chemicals were purchased from a variety of manufacturers including; Aldrich Chemical Co., BDH Pty. Ltd., Oxoid Pty. Ltd., Sigma Chemical Company and United States Biochemical Corp. Ampicillin, nuclease free BSA, deoxyribonucleotides, IPTG and X-Gal were purchased from Boehringer-Mannheim. Acrylamide, methylenebisacrylamide and caesium chloride were purchased from BDH. Agarose (type I) was purchased from the Sigma Chemical Co.

Radioactive isotopes

Alpha labelled ^{32}P 3' deoxycytidine triphosphate and alpha labelled ^{32}P 3' deoxyadenosine triphosphate (3000 μCi/mmol) were purchased from Bresatec Pty. Ltd. ^{35}S labelled thiourea was purchased from Amersham Pty. Ltd.

Nucleic acids

Dephosphorylated, EcoRI restricted bacteriophage lambda gt10 DNA arms and EcoRI adaptors were purchased from Promega Corporation and used to construct a cDNA library. Wild type bacteriophage lambda genomic DNA was purchased from Bresatec Pty. Ltd. and restricted with HindIII for use as a molecular weight marker. A number of plasmid clones were also used in this study and are presented in table 2.1.

Enzymes

Restriction endonucleases, proteinase K and pronase were purchased from Boehringer-Mannheim. Ribonuclease A and lysozyme were purchased from Sigma Chemical Co. Taq DNA polymerase was purchased from Bresatec Pty. Ltd. while bacteriophage T4 DNA ligase was purchased from New England Biolabs.

Table 2.1 : Plasmid clones

Plasmid	Insert	Vector	Reference
pSocE48	Spinach chloroplast large subunit of ribulose 1, 5-bisphosphate carboxylase	pBR325	Zurawski*et al*(1981)
pSoc7.7	spinach chloroplast 7.7 kbp PstI DNA fragment	pBR322	Palmer *et al* (1981)
pNH1	Spinach nuclear DNA fragment containing a 0.6 kbp Bam-EcoRI fragment with 99% homology to the spinach ct *psa*A gene	pBR322	Cheung and Scott (unpublished result)
FB7-12	Tobacco chitinase cDNA clone	pUC118	Neale *et al* (1990)
FB7-13	Tobacco chitinase cDNA clone	pGEM3z	Neale *et al* (1990)
pCU18	25s, 18s and transcribed spacer of the cucumber rRNA genes	pACYC184	Kavanagh and Timmis(1986)
pTBa1	19.6 kbp of the tobacco chloroplast genome	pBR322	Sugiura *et al* (1986)
pTBa2	18.1 kbp "	"	"
pTBa5	7.1 kbp "	"	"
pTB7	10.9 kbp "	"	"
pTB13	10.4 kbp "	"	"
pTB18	9.9 kbp "	"	"
pTB19	12.3 kbp "	"	"
pTB20	17.3 kbp "	"	"
pTB22	4.8 kbp "	"	"
pTB25	16.0 kbp "	"	"
pTB28	7.3 kbp "	"	"
pTB29	3.4 kbp "	"	"
pTBX6	9.6 kbp "	"	"

Plant Material Utilized In This Study

species	location
Arabidopsis thaliana	chapter 3
Beta vulgaris var. silverbeet	chapter 3 and 7
var. crimson globe	chapter 3
Chenopodium quinoa	chapter 3
Gossypium hirsutum var. deltapine 90	chapter 3
var. sicala	chapter 3
var. siokra	chapter 3
Hordeum vulgare var. Swift	chapter 3
var. Cleremont-France	chapter 3
var. Hiproly	chapter 3
var. India 115	chapter 3
Nicotiana tabacum var. W38	chapters 3-7
Pisum sativum	chapter 3
Tulbaghia violacea	chapter 3

1 x reaction buffers

Restriction endonuclease buffers were supplied with the enzymes when purchased from the manufacturer.

T4 DNA ligase ligation buffer: 10mM $MgCl_2$, 10mM DTT, 1mM spermidine, 1mM ATP, 50mM Tris-HCl (pH 7.5).

Taq DNA polymerase buffer: 67mM Tris-HCl (pH 8.8), 16.6mM $(NH_4)SO_4$, 0.45% Triton X-100, 200 µg/ml gelatin.

2.2: **Stock solutions**

1 x Denhardts Reagent: 0.02% (w/v) Ficoll, 0.02% (w/v) gelatin, 0.02% (w/v) polyvinylpyrrolidone.

1 x load buffer for agarose gels: 0.042% bromophenol blue, 6.67% (w/v) sucrose.

1 x load buffer for denaturing polyacrylamide gels: 95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol.

SM buffer: 0.1M NaCl, 50mM Tris-HCl (pH 7.5), 8mM $MgSO_4$, 0.1% gelatin.

Phenol/chloroform/isoamyl alcohol: consisted of 50% (v/v) phenol extracted with 1M Tris (pH 8), 2% β-mercaptoethanol, 0.1% hydroxyquinoline; 48% (v/v) chloroform; 2% (v/v) isoamyl alcohol.

1 x SSC: 0.15M NaCl, 0.15M $Na_3C_6H_50_7.2H_20$, pH7.2.

1 x SSPE: 0.1M NaCl, 10mM NaH_2PO_4, 1mM EDTA, pH 7.4.

1 x TAE: 40mM Tris, 20mM $NaCH_3COO$, 2mM EDTA, pH 7.8.

1 x TBE: 100mM Tris, 89mM boric acid, 2mM EDTA, pH 8.3

1 x TE: 10mM Tris-HCl (pH 8), 1mM EDTA.

2.3: **Bacterial growth media**

Luria broth: 1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, pH 7.2.

Luria agar/agarose plates: 1.5% (w/v) agar or agarose in Luria broth.

Ampicillin plates: 1.5% Luria agar plates with 50 µg/ml of ampicillin.

X-Gal plates: Ampicillin plates with a 3 ml Luria broth overlay containing 0.8% agar, 0.03% (w/v) X-gal, 0.3mM IPTG and 50 µg/ml of ampicillin.

2.4: ***Escherichia coli* and lambda bacteriophage strains**

Strains of *E.coli* and lambda that were used are presented in table 2.2.

Methods

2.5: **Isolation of plasmid and bacteriophage DNA**

Large scale plasmid DNA isolation

E. coli cells were grown at 37°C in 500 mls of Luria broth containing an appropriate antibiotic, until an $O.D._{600}$ of 0.6 was achieved. Chloramphenicol (170 µg/ml) was then added to the culture and incubation was continued for another 14 to 16 hours. Bacterial cells were harvested by centrifugation (5000 x g, for 10 minutes) and resuspended in 10 mls of 50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH 8.0), 2 mg/ml of lysozyme and placed on ice. 20 mls of 0.2M NaOH, 1% SDS was gently mixed into the bacterial suspension which was then stood on ice for 5 minutes. 15 mls

Table 2.2 : *E. coli* and lambda strains

Strain	mcrA	mcrB	Use	Reference
E. coli				
NW2	-	-	Host for propagation of lambda EMBL4 tobacco genomic library	Woodcock *et al* (1988)
NM538	-	+	Assay and propagation of lambda EMBL4	Raleigh *et al* (1988)
NM539	-	+	*Spi* selection of lambda EMBL4	Frischauff *et al* (1983)
JM101	+	?	Host for pUC19 cloning	Messing (1979)
C600	-	+	Titering and propagation of lambda gt10	Huynh *et al* (1985)
C600 Hfl	?	?	Selection of recombinant lambda gt10 and cDNA library amplification	Hunyh *et al* (1985)
Lambda				
EMBL4			Construction of a tobacco nuclear genomic library	Frischauff *et al* (1983)
gt10			Construction of a tobacco cDNA library	Huynh *et al* 1985)

of 3M NaOAc (pH 4.8) was added and the suspension left on ice for another 60 minutes. The bacterial lysate was centrifuged at 17000 x g for 15 minutes and nucleic acids remaining in the supernatant, ethanol precipitated and dissolved in TE. Plasmid DNA was finally purified on a caesium chloride/ethidium bromide gradient (Maniatis *et al*, 1982).

Small scale plasmid DNA isolation

Small amounts of plasmid DNA were extracted from 10 ml overnight *E. coli* cultures by the alkaline lysis method of Sambrook *et al* (1989).

Isolation of lambda bacteriophage DNA

Lambda DNA was prepared by a modification of the plate lysate method of Maniatis *et al* (1982). Lambda bacteriophage were plated with an appropriate *E. coli* host onto agarose plates such that confluent lysis was achieved. SM buffer (3 ml) was washed across the surface of the plate for 3 hours at room temperature and the buffer then recollected. An equal volume of Whatman DEAE cellulose suspended in 50mM Tris-HCl (pH 7.2), 10mM $MgCl_2$ was added and the mixture shaken vigorously for ten minutes at room temperature. Bacterial debris and DEAE cellulose were removed by centrifugation (4000 x g for 10 minutes) and the supernatant incubated with 1 mg/ml of DNase I and RNase A for 30 minutes at 37°C. Bacteriophage were precipitated by the addition of an equal volume of 20% (w/v) polyethlyene glycol, 2M NaCl in SM buffer and centrifugation at 8000 x g for 20 minutes at 4°C. The pellet was resuspended in 0.5 mls of SM buffer and the phage particles disrupted by the addition of 5 µl of 10% SDS and 5 µl of 0.5M EDTA (pH 8.0). Following a phenol/chloroform/isoamyl alcohol extraction the bacteriophage DNA was precipitated with an equal volume of isopropanol.

2.6: Isolation of plant nucleic acids

Large scale isolation of total leaf DNA

Total DNA was isolated from plant leaf tissue as described by Scott and Possingham (1980). Fresh leaf tissue (5-10g) was frozen in liquid nitrogen, ground to fine powder in a mortar and pestle and thawed in several volumes of 2 x SSC, 50mM EDTA, 1% sarkosyl. This extract was incubated with RNase A and pronase for 1 hour at 37°C and the supernatant collected after centrifugation at 12000 x g for 15 minutes. Nucleic acids were then spun out of solution by centrifugation at 60000 r.p.m. for three hours in a Beckman type 70Ti rotor. The resultant pellet was dissolved in TE and further purified by caesium chloride gradient centrifugation.

Small scale isolation of total leaf DNA

When leaf tissue was in limited supply 200 mg of leaf tissue was frozen in liquid nitrogen, ground to a fine powder and thawed in 1 ml of 0.1M Tris-HCl (pH 7.2), 10mM EDTA, 2% sarkosyl. This suspension was incubated at 65°C for 15 minutes followed by a single phenol/chloroform/isoamyl alcohol extraction. The aqueous phase was then made up to 1.5M NaCl and 200 µl loaded onto a 1 ml Sepharose CL-4B column (Sigma Chemical Company) equilabrated with 1.5M NaCl, 0.1M Tris-HCl (pH 7.2), 10mM EDTA. The first 500 µl of eluate was discarded and DNA collected in the next 700 µl fraction. Nucleic acids were then precipitated with isopropanol and resuspended in TE.

Isolation of tobacco nuclear DNA

Tobacco nuclear DNA was isolated from leaf tissue by the method of van Loon *et al* (1975). Leaf tissue was homogenized in six volumes of ice cold 0.5M Tris-HCl (pH 7.6), 0.3M sucrose, 5mM $MgCl_2$, 0.5% β-mercaptoethanol and the homogenate filtered twice through miracloth. The filtrate was made up to 2% Triton X-100 and nuclei pelleted by centrifugation (270 x g for 10 minutes at 4°C). Upon resuspension in homogenization buffer plus Triton X-100 nuclei were again pelleted by centrifugation

and resuspended in 10 mls of homogenization buffer. This suspension was centrifuged at 4000 x g for 30 minutes at $4^{o}C$ through a 15 ml cushion of 0.01M Tris-HCl (pH 7.6), 1.78M sucrose, 5mM $MgCl_2$, 5mM β-mercaptoethanol. The nuclei pellet was then lysed in 2 x SSC, 50mM EDTA, 1% sarkosyl and the lysate phenol/chloroform/isoamyl alcohol extracted followed by precipitation of nucleic acids with ethanol.

Isolation of tobacco mitochondrial DNA

Fresh tobacco leaf tissue was chilled on ice prior to homogenization in 0.3M sucrose, 25mM tetrasodium pyrophosphate, 20mM ascorbic acid, 10mM KH_2PO_4, 2mM EDTA, 1% polyvinylpyrrolidone, 1% BSA, pH 7.5. The crude lysate was filtered through miracloth and then centrifuged at 700 x g for 5 minutes at $4^{o}C$. The supernatant was retained and centrifuged at 11000 x g for 20 minutes at $4^{o}C$ to pellet mitochondria. The mitochondria were resuspended in homogenization buffer and pelleted by centrifugation several times before being spun through a Percoll gradient. The purified mitochondria were then treated with DNAse grade I to remove any contaminating genomic or plastid DNA. Mitochondria were lysed in 2 x SSC, 50mM EDTA, 1% sarkosyl and the lysate phenol/chloroform/isoamyl alcohol extracted before precipitation of nucleic acids with ethanol.

Isolation of plant RNA

Total RNA was isolated from leaf tissue by a modification of the phenol/SDS method of Ausubel *et al* (1987). Leaf tissue (15g) was frozen in liquid nitrogen, ground to a fine powder in a mortar and pestle and homogenized in 150 mls of 0.18M Tris-HCl (pH 8.2), 0.09M NaCl, 4.5mM EDTA, 1% SDS with 50 mls of phenol equilibrated in 0.2M Tris-HCl (pH 8.2), 0.1M NaCl, 5mM EDTA. Chloroform (50mls) was added and the extract homogenized briefly. The resultant slurry was heated for 20 minutes at $50^{o}C$ and then centrifuged at 11000 x g for 20 minutes at $4^{o}C$. The aqueous phase was re-extracted several times with equal volumes of phenol and chloroform, followed by ethanol precipitation of nucleic acids.

The nucleic acids were dissolved in 2 mls of dH_2O and RNA differentially precipitated from DNA by the addition of 6 mls of ice cold 4M NaAc, 10mM EDTA, (pH 6) and centrifugation at 17000 x g for 20 minutes at 0^oC. The RNA pellet was resuspended in dH_2O and the RNA reprecipitated with two volumes of ethanol.

Isolation of polyadenylated mRNA

Polyadenylated mRNA was purified from total plant RNA using Hybond-mAP (messenger affinity paper) as per manufacturers instructions (Amersham).

2.7: Recombinant DNA techniques

General recombinant DNA techniques such as restriction enzyme digestions, DNA subcloning etc. were done according to standard protocols (Sambrook *et al*, 1989) or as recommended by biochemical manufacturers.

2.8: Electrophoresis

DNA agarose gel electrophoresis

Agarose gel electrophoresis of DNA was performed in submarine gel tanks with 1 x TAE running buffer and gels made of 1 x TAE and an appropriate concentration of agarose. DNA samples were loaded in 0.04% bromophenol blue, 6.7% (w/v) sucrose and bacteriophage lambda DNA restricted with HindIII used as a molecular weight marker. DNA was recovered from agarose gels by the freeze squeeze method of Thuring *et al* (1975).

RNA denaturing agarose gel electrophoresis

RNA was electrophoresed in gels consisting of 1.5% (w/v) agarose, 10mM sodium orthophosphate (pH 7.0), 8% formaldehyde with 10mM sodium orthophosphate (pH 7.0) used as a running buffer. RNA samples were denatured prior to electrophoresis by heating at 65^oC for 10 minutes in 50% (v/v) formamide, 12% (v/v) formaldehyde, 10mM sodium orthophosphate (pH 7.0).

Denaturing polyacrylamide gel electrophoresis

DNA sequencing reactions were electrophoresed at 1500 volts in a gel consisting of 5% acrylamide, 7M urea, 1 x TBE using a BRL Sequencing Gel Electrophoresis System, Model S2 (BRL).

2.9: Nucleic acid membrane transfer and hybridization conditions

Southern blotting

Following electrophoresis, agarose gels were denatured for 30 minutes in 0.8M NaCl, 0.4M NaOH and neutralized for 45 minutes in 1.5M NaCl, 0.5M Tris-HCl (pH 7.2). DNA was then transferred to a nitrocellulose membrane (Schleicher and Schuell) by the method of Southern *et al* (1975).

Dot blotting

Dot blotting of DNA samples was done using a Minifold II Micro-Sample Filtration Manifold (Schleicher and Schuell). DNA samples were denatured with an equal volume of NaOH for 10 minutes at room temperature, then cooled on ice and neutralized with 1 volume of ice cold 0.25M HCl, 0.25M Tris (pH 8.0), 1.5M NaCl prior to dot blotting onto a nitrocellulose membrane.

Plaque lifts

Plaque lifting to nitrocellulose was done as described by Maniatis *et al* (1982). Bacteriophage were plated onto agar plates with an appropriate *E. coli* host in an agarose overlay. Bacteriophage DNA was transferred to a nitrocellulose filter where it was denatured in 1.5M NaCl, 0.5M NaOH for 30 seconds and then neutralized in 1.5M NaCl, 0.5M Tris-HCl (pH 8.0) for 5 minutes. Plaque lifts were duplicated for each plate.

Following either dot blotting, plaque lifting or Southern transfer, DNA was fixed to the nitrocellulose membrane by baking at 80°C *in vacuo* for 3 hours.

Southern blot, dot blot and plaque lift membrane hybridization conditions

Membranes were hybridized with ^{32}P labelled DNA probes in a hybridization solution of 4 x SSC, 5 x Denhardts reagent, 0.1% SDS and with 100 µg of denatured, sonicated salmon testes DNA per ml of hybridization solution. Hybridization was conducted at 65°C for 14 to 16 hours. Membranes were then washed 3 times in 2 x SSC, 0.1% SDS at 65°C for 30 minutes. Autoradiography was carried out using intensifying screens except when autoradiographs were used for quantitative analysis.

Northern blotting

Denaturing formamide agarose gels were soaked for two hours in a large volume of dH_2O after electrophoresis of RNA samples. The RNA was then blotted overnight onto a nitrocellulose membrane using 20 x SSC as a transfer solvent. The membrane was subsequently baked at 80°C *in vacuo* for 3 hours.

Northern blot hybridization conditions

Northern blot membranes were hybridized at 42°C for 14 to 16 hours in a hybridization solution of 50% formamide, 2 x Denhardts reagent, 0.1% SDS, 5 x SSPE and 100 µg of denatured, sonicated salmon testes DNA per ml of hybridization solution. Following hybridization membranes were sequentially washed for 15 minutes at room temperature in 2 x SSC, 0.1% SDS; 0.5 x SSC, 0.1% SDS and 0.1 x SSC, 0.1% SDS. Non specific binding of the probe onto the membrane was further reduced by washing in 2 x SSC, 0.1% SDS at 65°C for 15 minutes if required. Autoradiography was carried out using intensifying screens.

Radioactive labelling of DNA probes for hybridization

Routinely 50 ng of DNA was labelled using a Bresatec Oligo-labelling kit with 25 µCi of alpha labelled ^{32}P deoxyadenosine 5'triphosphate (Bresatec). Unincorporated nucleotides were separated from labelled DNA by passage through a column of Bio-Gel P-60 polyacrylamide beads (Bio-Rad). DNA probes were denatured in 0.2M NaOH for 10 minutes at room temperature and neutralized with 0.2M HCl just prior to addition to

the hybridization solution. Cloned DNA fragments used as hybridization probes were always separated from vector sequences by agarose gel electrophoresis. This process was repeated twice for DNA fragments used as probes for dot blot quantitation.

2.10: Double stranded plasmid DNA sequencing

Preparation of template DNA

DNA fragments to be sequenced were subcloned into pUC19 and plasmid DNA suitable for sequencing prepared by the alkaline lysis miniprep method of Sambrook *et al* (1989). DNA template (3 μg) was RNase treated before being denatured in 0.2M NaOH, 0.2mM EDTA at 37°C for 15 minutes and then spun through a Sepharose CL-6B column (Sigma Chemical Co.). 1 μg of this denatured template was then used per sequencing reaction.

DNA sequencing reactions

DNA templates were sequenced by the dideoxy chain termination method of Sanger *et al* (1977) using alpha labelled ^{32}P deoxyadenosine 5'-triphosphate and a Sequenase Version 2.0 kit (United States Biochemical Corp.) which contains a modified T7 DNA polymerase. Sequencing reactions were electrophoresed on denaturing polyacrylamide gels (see Methods Electrophoresis). Gels were fixed in 20% (v/v) methanol, 10% (v/v) acetic acid for 30 minutes, dried for 2 hours at 65°C and exposed to X-ray film overnight. DNA sequencing was done using either an overlapping set of subclones of the DNA fragment to be sequenced or specifically synthesized oligonucleotide primers, or a combination of both. All DNA sequence data presented has been bidirectionally sequenced from overlapping sets of subclones. Sequence data was analysed using Staden, NIH programs and the Genetics Computer Group Sequence Analysis Software Package Version 6.1 (Devereux *et al* 1984).

2.11: **Oligonucleotide primer synthesis**

Oligonucleotides were synthesized on 0.2 mmol CPG columns using a PCR-MATE DNA synthesizer (Applied Biosystems). Nucleotide deprotection by removal of dimethoxytrityl was monitored by determining the $O.D._{498}$ of the TCA and acetonitrile eluate from the column. Following synthesis, oligonucleotides were eluted from the CPG column by treating with fresh, concentrated ammonium hydroxide for 1 hour. Benzoyl and isobutyryl base protecting groups were removed by incubating the solution at $55^{o}C$ for 8 to 15 hours. By-products and failure sequences were removed by applying the ammonia solution to an OPC cartridge. Purified oligonucleotides were eluted from the cartridge with 1 ml of 20% acetonitrile, desiccated and resuspended in 1xTE.

Some oligonucleotides were also purchased from the Department of Microbiology and Immunology, University of Adelaide.

2.12: **Polymerase chain reactions**

Polymerase chain reactions were carried out as described by Sambrook *et al* (1989). Genomic DNA (1 µg) underwent PCR in 1 x Taq DNA polymerase buffer, 1.5mM $MgCl_2$, 1pmol of each primer, 1.25mM of each deoxynucleoside triphosphate and 2.5 units of Taq DNA polymerase. Reaction mixtures were overlaid with paraffin oil to prevent sample evaporation. Thermal cycling was carried out for 30 cycles under the temperatures recommended by Sambrook *et al* (1989).

CHAPTER 3

Plastid DNA Sequence Homologies Within Higher Plant Nuclear Genomes

Plastid DNA sequence homologies were first identified within the spinach nuclear genome by Southern blot analysis (Timmis and Steele Scott, 1983). The nuclear genome of many higher plants is extensively methylated at cytosine residues (Doerfler, 1990; Grenbaum *et al*, 1981) whereas the chloroplast DNA of these species generally appears to be unmethylated (Smillie and Scott, 1969; Scott and Possingham, 1980). These distinct methylation states have been exploited to allow the separation of plant nuclear DNA from plastid DNA by differential digestion with methylation sensitive restriction enzymes. Of the few reports of plastid DNA methylation in higher plants all have shown that where methylation occurs it involves very few sites within the plastid genome (Ngernprasirtsiri, 1988a, 1988b, 1989). Consequently a majority of the genome remains susceptible to cleavage by methylation sensitive restriction enzymes.

Restriction of total spinach DNA with HpaII results in cleavage of the unmethylated plastid genome while the extensively methylated nuclear genome remains essentially intact. HpaII is a methylation sensitive restriction enzyme that recognizes the sequence CCGG but will not restrict when the internal cytosine is methylated. Hybridization of NH1, a 0.6 kbp BamHI-EcoRI DNA fragment with 99% homology to the spinach plastid *psa*A gene and 5' flanking regions (Cheung and Steele Scott, unpublished results), to HpaII restricted total spinach DNA is shown in figure 3.1, track 1. Four distinct low molecular weight bands are evident in addition to a high molecular weight smear of DNA with probe homology. These four distinct bands are clearly of plastid origin (figure 3.1, track 2) while the high molecular weight smear represents apparently unrestricted nuclear DNA with homology to the plastid probe. A majority of the homology of the spinach plastid genome to this probe is present as a 1.2 kbp HpaII fragment, which presumably contains part of the spinach *psa*A gene. The remaining less prominent bands of 1.7 kbp, 2.5 kbp and 3 kbp that are clearly plastid in origin (figure

3.1, track 2) most likely represent other smaller regions of homology between the plastid genome and this probe.

Upon digestion of total spinach leaf DNA with HpaII plus EcoRI the high molecular weight nuclear smear present in figure 3.1, track 1 is reduced to at least seven discrete EcoRI fragments (figure 3.1, track 4). The origin of these fragments is undoubtly nuclear or mitochondrial as they are not present in the same restriction of plastid DNA (figure 3.1, track 3). These results obtained with total leaf DNA extracts exactly parallel those obtained using purified nuclear and chloroplast DNA preparations (Steele Scott and Timmis, 1984; Timmis and Steele Scott, 1983).

Southern hybridization has identified sequence homologies between virtually the entire plastome and the spinach nuclear genome (Timmis and Steele Scott, 1983; Steele Scott and Timmis, 1984). These homologies within the spinach nuclear genome appear to be less than 2-3 kbp in length and it was estimated that five copies of the plastome were present within a haploid spinach nuclear genome (Steele Scott and Timmis, 1984). However this estimate was based on hybridization of plastid sequences to nuclear integrants with potentially variable homology to these probes. Consequently the absolute copy number of plastid homologies may have been underestimated.

It would appear that plastid DNA sequence homologies are a common phenomenon in higher plant nuclear genomes and may involve the entire plastid genome. However little is known about the conservation or stablity of these sequences within a plant species. As plastid sequences within the nuclear genome are presumed to be predominantly nonfunctional they may be frequently rearranged or deleted from the genome, thereby showing a high level of hypervariability within or between plant species. The following chapter describes an examination of the nature and stability of promiscuous plastid sequences present within the nuclear genomes of individuals from a number of different plant species.

3.1: Species With A Constant Nuclear Hybridization Pattern To Plastid DNA Probes

For a number of plant species, individual members within the species showed identical hybridization patterns to plastid DNA probes. Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to cotton (*Gossypium hirsutum*) total leaf DNA restricted with HpaII + EcoRI produced a spectrum of higher molecular weight nuclear fragments with homology to this probe. Approximately a dozen homologies were visible, superimposed on a heterogeneous background of less homologous fragments (figure 3.2, tracks 1-6). These latter homologies represent either smaller plastid sequences integrated into the nuclear genome or indicate the presence of sequences with reduced homology to this probe. An additional series of strongly hybridizing low molecular weight fragments (ie. $<$ 2kbp) are visible which represent HpaII restriction products of the plastid genome. The greater intensity of hybridization of these small fragments is due to the high multiplicity of the plastid genome within an individual plant cell.

No variation in hybridization pattern is observed between individual plants of either the same or a different cotton variety (figure 3.2). An identical hybridization pattern was also obtained for plant members of a third cotton variety, *siokra* (result not shown), suggesting apparent homogeneity of these sequences within this species.

Chenopodium quinoa displayed similar homogeneity of nuclear sequences with homology to NH1, (figure 3.3). In this species a dozen bands representing nuclear sequences with homology to this probe were identifiable in addition to a series of smaller, more intense bands representing *bona fide* plastid DNA homology. These nuclear and plastid sequences were identical in three individual plant DNAs examined.

Isolated DNA from four individual *Nicotiana tabacum* plants also showed identical spectra of nuclear EcoRI restriction fragments with homology to the spinach *rbc*L gene (figure 3.4a) and to the spinach plastid 7.7kbp PstI DNA probe (figure 3.4b). In addition to a series of minor nuclear homologies, both plastid probes identifed a methylated EcoRI restriction fragment that was identical in size to the *bona fide* plastid

EcoRI restriction fragment observed in an EcoRI digestion of total DNA (figure 3.4a and 4b, track 5). The apparent conservation of restriction sites shown by these relatively strongly hybridizing nuclear sequences suggests that large tracts of plastid DNA may be integrated into the tobacco nuclear genome. This result parallels observations in *Chenopodium album, Atriplex cinerea* (Ayliffe *et al* , 1987) and potato (du Jardin, 1990). The nature of these tobacco nuclear sequences was examined in more detail and is described in chapter four.

All individuals from these three plant species show a species specific hybridization pattern to plastid DNA probes. However the number of individual plants examined for each species was relatively small and a more extensive analysis may identify some intraspecies variability.

3.2: **Species Which Show Variation In Nuclear Hybridization Patterns To Plastid DNA Probes**

In contrast to the uniformity of promiscuous plastid sequence homologies observed for three different cotton varieties, several other agricultural species showed a cultivar specific spectrum of plastid DNA homologies in the nuclear genome. Hybridization of NH1 to HpaII + EcoRI restricted total leaf DNA from pea (*Pisum sativum*) revealed relatively few nuclear sequences homologous to this probe (figure 3.5). However the arrangement of these nuclear sequences was unique for all nine lines of pea that were examined (figure 3.5).

Four cultivars of barley that were examined each contained a unique and distinguishing spectrum of high molecular weight (>2 kbp) fragments of presumed nuclear origin that were homologous to the spinach plastid 7.7 kbp PstI DNA fragment (figure 3.6). Hybridization of the same probe to DNA isolated from several individual plants of three barley varieties showed no evidence of variation in the spectrum of nuclear homologies within a cultivar (figure 3.7). Nuclear DNA from cultivars of wheat also possessed a unique hybridization pattern when restricted and probed with plastid

DNA and again no variation in this pattern was observed between individual members of the same cultivar (results not shown).

It appears that these plastid DNA probes may therefore be used to distinguish varieties of certain commercial crops that can not be easily indentified morphologically. The major criterion for a useful plant DNA fingerprinting probe is that it must be capable of distinguishing between different plant species in addition to different varieties of the same species. The DNA hybridization pattern of any individual within a plant variety should be identical, as should the hybridization pattern of DNA extracted from different tissues of the same plant. A universal probe capable of DNA fingerprinting a wide range of species would be advantageous. For some plant species such as barley, wheat and pea, plastid sequence homologies within the nuclear genome may act as suitable target sequences for DNA fingerprint analysis. The almost ubiquitous nature of these sequences within higher plants suggests that other plant cultivars may also possess unique spectrums of these homologies.

However, as target sequences for fingerprinting probes, nuclear sequences with plastid DNA homology have several disadvantages. Techniqually it is difficult to show these nuclear sequence homologies by Southern hybridization as they generally appear low in copy number and are maintained in large plant genomes. Double restriction enzyme digestion and large agarose gels are required to separate *bona fide* plastid sequences from these nuclear homologies, which is both time consuming, expensive and increases the likelihood of partial restriction products. Finally plastid sequence homologies are also found in mitochondrial DNA, potentially confusing fingerprint patterns due to uniparental inheritance of this organellar genome.

3.3: **Individual Plant Variation In Nuclear Hybridization Patterns To Plastid DNA Probes**

Hybridization of NH1 to HpaII + EcoRI restricted DNA from six spinach plants (figure 3.8) identifies three major homologies of *bona fide* plastid origin (compare with figure 3.1, track 3) in addition to a series of discrete EcoRI restriction fragments

representing nuclear homologies to this plastid probe. Four of these samples show an identical spectrum of nuclear homologies (figure 3.8, tracks 2, 3, 4, 6) while the remaining two DNAs each possess a unique arrangement. A 5.5 kbp band is absent in one of these individuals (figure 3.8, track 1) while an additional 9 kbp band is present in the other (figure 3.8, track 5). A series of faint nuclear homologies are present in these six plant DNAs that were not previously identified (figure 3.1, track 4) reflecting the higher sensitivity of this particular experiment.

Therefore a degree of variation in these sequences is observed amongst individual plants of this variety. These differences could be accounted for by numerous events including heterozygosity and allelic variation, deletion, duplication, endopolyploidy, differential methylation, base substitutions creating or destroying restriction sites, mitochondrial polymorphisms or a new insertion of a plastid sequence within the nuclear genome.

3.4: Hypervariability Of Plastid DNA Sequence Homologies In The Nuclear Genome Of *Beta vulgaris*

Initial characterization of plastid DNA sequence homologies in the nuclear genome of *Beta vulgaris* revealed heterogenity of these sequences amongst different cultivars of the species. Hybridization of NH1 to HpaII + EcoRI restricted DNA from two *B. vulgaris* cultivars (crimson globe and silverbeet) identified many nuclear homologies to this plastid DNA probe in both varieties (figure 3.9). These nuclear homologies were often to bands of common molecular weight in both cultivars, although clear differences in the hybridization pattern of both varieties was evident. There were also some discrete bands produced by HpaII digestion which were different for each variety. This is consistent with earlier observations of a small number of homologies to plastid DNA in regions of the spinach nuclear genome with reduced DNA methylation (Steele Scott and Timmis, 1984).

Analysis of a number of DNAs extracted from individual plants of both of these cultivars revealed a suprisingly large amount of heterogenity. A unique spectrum of

homologies was observed for each plant when NH1 was hybridized to HpaII + EcoRI restricted DNAs isolated from seven individual *B. vulgaris* var. silverbeet plants (figure 3.10a). Many hybridizing fragments were common to more than one plant, but each plant possessed additional bands that distinguished it from every other plant.

In addition to this variation in hybridization pattern, quantitative differences may exist between plants. It is likely that bands common to two plant DNAs represent the same plastid DNA insertion within the nuclear genome as the sequences comigrate and are homologous to the same probe. DNAs in figure 3.10a, track 3 and 4 show nuclear homologies to 12.8 kbp and 8.5 kbp bands, both of which are marked with a cross. Laser densitometery showed that the 12.8 kbp fragment in track 4 has a hybridization intensity three times that of the equivalent band in track 3, while both DNAs contain approximately equivalent ratios of the 8.5 kbp homology. This difference can not therefore be ascribed to DNA loading differences and thus is likely to represent a quantitative difference between one of the sequences common to these two individual plant DNAs. Several other analogous differences are also evident between other plant DNAs. These apparent quantifiable differences could be the result of either deletion or amplification of pre-existing nuclear homologies or a copy number difference arising from heterozygo sity at specific promiscuous loci.

Hybridization of the same Southern blot with the spinach *rbc*L gene produced a similiar result. Each plant again possessed a unique spectrum of higher molecular weight nuclear sequences with homology to this plastid probe (figure 3.10b) of which many were again common to more than one plant.

Likewise DNAs from eight individual *B. vulgaris* var. crimson globe plants each contained a unique restriction pattern of nuclear sequences that were homologous to either NH1 (figure 3.11a) or to the spinach *rbc*L gene (figure 3.11b). To ensure that the variation observed between DNAs of individual plants of both *B. vulgaris* varieties was not the result of partial restriction, digestions were incubated for excess time with high concentrations of restriction enzyme. In addition the same Southern filters were rehybridized with a probe (pCU18) encoding a complete cucumber rRNA gene repeat

unit (Kavanagh and Timmis, 1986). Plant rRNA genes are tandemly repeated throughout the plant genome (Kavanagh and Timmis, 1986). Any evidence of partial restriction should therefore be readily observed with this probe. Each plant DNA produced an identical hybridization pattern to this probe (figure 3.10c and 3.11c) suggesting that partial restriction was not present.

3.5: **Breeding Systems**

It therefore appears that within the higher plant kingdom promiscuous plastid sequences are homogeneous within the nuclear genomes of some plant species but remarkably heterogeneous in others. In *B. vulgaris* as great or greater variation in these sequences is observed between individual plants of the same variety when compared to plants from a different cultivar. Conversely morphologically diverse varieties of cotton show apparent genetic uniformity of these promiscuous sequences.

Only two plant species demonstrated variability of these nuclear sequences of plastid origin at an individual plant level, these being *S. oleracea* and *B. vulgaris*. Spinach is an obligate outbreeder and is dioecious (Allard, 1960; Richards, 1986; Simmonds, 1976) while outcrossing normally occurs for beet which has a genetic incompatability system (Lunquist *et al*, 1973). Conversely the other plant species examined are all self pollinating crop plants (Allard, 1960; Simmonds, 1976, 1971). The sequence variability observed in spinach and beet may therefore reflect the diversity of the gene pool from which these species were derived and this diversity is maintained by continual outbreeding. The remaining plants, most of which are crop plants, have therefore tended towards genetic homogeneity by continual self-fertilization.

Cultivar variation of these nuclear homologies is observed in barley but absent in three lines of cotton examined although both of these species employ inbreeding reproductive mechanisms. This contrasting result probably reflects the age of divergence of the different plant lines in each of these species and the diversity of the gene pool from which they were selected.

Given the widespread occurence of plastid sequence homologies in higher plant genomes, these sequences may serve as a useful index of genomic diversity within a plant population or species. Traditionally genetic diversity within a plant population is measured by determining the number of disease resistance genes, Mendelian morphological markers and isoenzyme variants within the population in addition to examing other economic and metric traits (Brown, 1978; Marshall and Allard, 1970; Hvids and Nielsen, 1977). These approaches consider only a very small portion of the entire genome and these regions are under functional constraints. In contrast promiscuous plastid sequences are probably dispersed throughout the entire plant genome [plastid homologies have been identified on chromosome 5, 9 and 10 in tomato (Pichersky and Tanksley, 1988; Pichersky *et al*, 1991)] and integrated into regions which appear to be in a more dynamic state of genomic flux. Analysis of these sequences could therefore assist in giving a more general estimation of genomic and genetic variability within a plant population.

3.6: **Somatic Variation Of Nuclear Sequences With Plastid DNA Homology**

Analysis of promiscuous plastid sequences in DNA extracted from different tissues of the same *B. vulgaris* plant, in some cases revealed heterogeneity of these sequences. Hybridization of NH1 to DNA extracted from leaf and petiole tissue of two silverbeet plants showed clear heterogeneity in the banding pattern observed for DNA extracted from different tissues of the same plant (figure 3.12a tracks 5, 7). As would be expected a majority of the nuclear homologies were common to all DNAs of the same plant, however four bands of 3.4, 3, 1.8 and 1.6 kbp were unique to leaf DNA (figure 3.12a, track 5) while two bands of 2.95 kbp and 2.76 kbp were found only in petiole tissue (figure 3.12a, track 7). Conversley no differences in hybridization pattern could be detected between DNAs isolated from the leaf and petiole tissue of a second plant (figure 3.12a, tracks 1 and 3).

Hybridization of the spinach *rbc*L gene to the same Southern filter again identified heterogeneity between the hybridization pattern of the same two DNAs. Two

bands of approximately 6.6 kbp were present in DNA extracted from petiole tissue (figure 3.12b, track 7) while only a single fragment of 6.6 kbp was hybridized in DNA extracted from leaf tissue (figure 3.12b, track 5). Hybridization of the ribosomal repeat (pCU18) to this filter did not identify any differences in banding patterns between rDNAs extracted from the same plant (figure 3.12c) suggesting that partial restriction was not the cause of the observed heterogeneity.

No differential hybridization patterns were observed in HpaII digestions of these DNAs (figure 3.12a and 3.12b, tracks 6 and 8) and these unique bands must therefore represent EcoRI restriction fragments. It is unlikely that the observed variation is the result of differing amounts of mitochondrial DNA being present in each tissue DNA preparation as both DNAs contain bands that are unique. A mitochondrial DNA polymorphism existing between different plant tissues could offer an expla nation for these results, however this is unlikely given the number of bands involved and their HpaII resistant nature. The number of different bands involved also suggests that base substitutions creating or destroying restriction sites are not the cause of these observations.

This heterogeneity could be the result of genomic rearrangement within somatic cells. These sequences are likely to be dispersed throughout the plant genome and their somatic variation suggests that they have integrated into hypervariable regions of the genome. Transfected sequences that integrate into mammalian genomes commonly show temporary instability and occasionally this instability is continuous (for review see Murnane, 1990). These continuously unstable locations undergo DNA amplification of both the integrated sequence and surrounding cell DNA. Prior to the integration event these genomic locations appear stable; that is the process of sequence integration can cause localized genomic instability. No evidence for sequence specificity has been observed (Murnane and Young, 1989). This instability may be a widespread phenomenon because large deletions in bacterial genomes can lead to subsequent genome instability resulting in gene amplification (Murnane, 1990).

At least one promiscuous plastid sequence has been shown to be amplified and deleted 10,000 fold during rice cell differentiation and dedifferentiation (Kikuchi *et al*, 1987) and flanking sequences were apparently involved. Two sequences present as dispersed repeats within both the nuclear and mitochondrial genomes of *Cucumis melo* are dramatically reduced in copy number during tissue culture. This copy number being only 10% in long term callus cultures as compared with the original plant material (Grisvard *et al*, 1990).

It is interesting to speculate that the difference in hybridization pattern of these DNAs may be the result of ongoing transposition of plastid sequences to the nucleus. If this is the case these results suggest that plastid sequences are being transposed to the nucleus at a very high frequency. In yeast it has been demonstrated that DNA can "escape" from the mitochondria to the nucleus at a frequency of 2×10^{-5} per cell per generation (Thorness and Fox, 1990). The nuclear background also affected the rate of mitochondrial DNA transfer to the yeast nuclear genome and by analogy could explain the observed heterogeneity of these sequences in some plants but not in others.

The somatic variability of these nuclear sequences with plastid DNA homology is not confined to *B. vulgaris* and has also been observed in some citrus species (Hughes, 1990).

3.7: Relationship between genome size and presence of plastid DNA homologies

Of the 15 plant species examined for nuclear homologies to plastid DNA sequences all contained some of these nuclear sequences with the exception of *Arabidopsis thaliana*. *A. thaliana* has a very small nuclear genome when compared with other plant species [1c= 0.2pg (Bennett and Smith, 1976)]. No sequence homology to the spinach plastid 7.7 kbp PstI DNA fragment could be detected in the nuclear genome of this species (figure 3.13, tracks 1-3), while nuclear homologies could be readily detected in an equivalent amount of DNA from *Tulbaghia violacea*, on the same Southern blot (figure 3.13, track 4). As the *A. thaliana* genome is 100 times smaller than that of *T. violacea* [1c= 19.8pg (Bennett and Smith, 1976)], approximately 100

times more nuclear genomes of *A. thaliana* were present on the filter and yet no nuclear homologies were detected to this plastid probe. This result suggests that promiscuous plastid sequences are located predominantly in nonessential regions of the nuclear genome, that are present in most plant species. In this respect it is notable that *A. thaliana* has far less repetitive nuclear DNA than most plants (Meyerowitz and Pruitt, 1985).

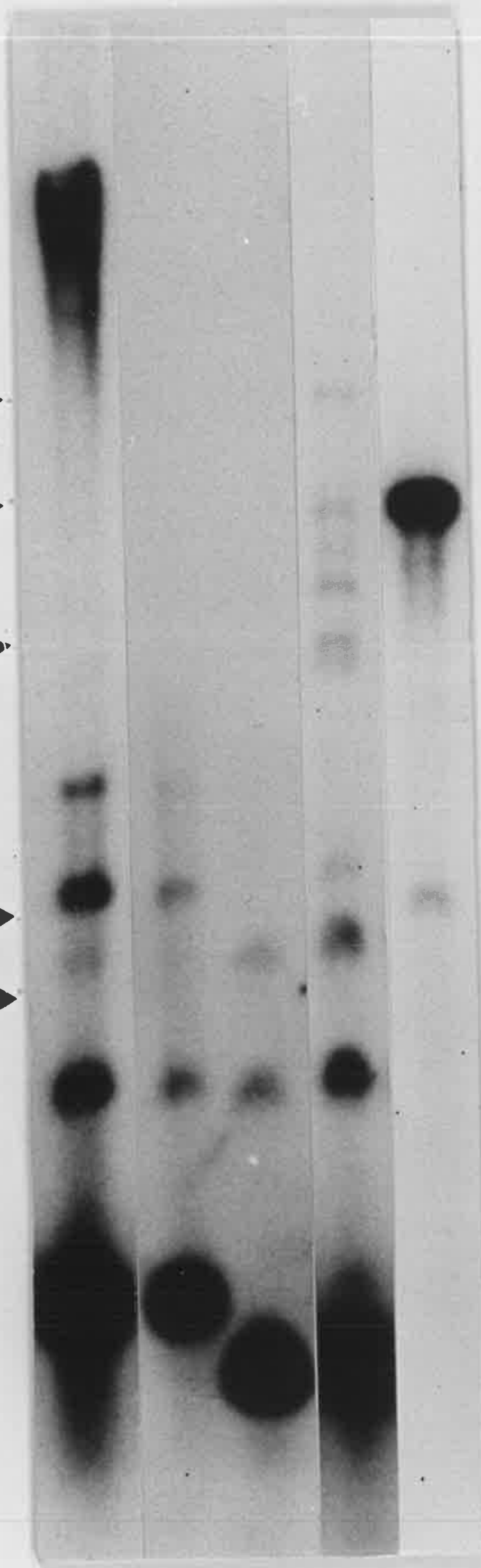
3.8: **Summary**

Plastid DNA sequence homologies are found within the nuclear genomes of a wide variety of higher plant species, including monocotyledonous and dicotyledonous species. In some species these sequences show apparent uniformity (eg. cotton, tobacco and *C. quinoa*) while in other species variability is observed. This variability ranges from differences detected between cultivars (eg. pea and barley) to the hypervariability observed in *B. vulgaris*. Somatic variation of these homologies in *B. vulgaris* suggests that these sequences either integrate into regions of the nuclear genome which rearrange at a high frequency or that a continual process of plastid sequence transposition to the plant nuclear genome is occurring. The abscence of promiscuous plastid sequences in the nuclear genome of *A. thaliana* suggests that a majority of these sequences are located in nonessential, repetitive regions of the plant genome. These plastid homologies may be suitable target sequences for DNA fingerprint analysis in some species, while for other plant species the variability of these sequences could serve as a measure of genomic heterogeneity within a population or species.

Figure 3.1: **Plastid DNA sequence homologies within the spinach (*Spinacea oleracea*) nuclear genome.**

Hybridization of NH1 to 0.5 μg of spinach plastid DNA (lanes 2, 3 and 5) and 5 μg of total leaf DNA (lanes 1 and 4). DNAs were restricted with: 1) and 2) HpaII, 3) and 4) HpaII + EcoRI and 5) EcoRI.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.



1 2 3 4 5

Figure 3.2: Plastid DNA sequence homologies within the nuclear genome of cotton (*Gossypium hirsutum*) show a constant hybridization pattern.

Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to cotton total leaf DNA. Each track contained 5 µg of DNA that had been isolated from a single plant and restricted with HpaII + EcoRI, except for lane 7 which contained 1 µg of DNA restricted with EcoRI. Lanes 1-3 contain DNAs isolated from individual plants of the cotton variety deltapine 90 while lanes 4-6 contained DNAs isolated from individual plants of the cotton variety sicala.

Plant members of a third cotton variety, siokra, produced a hybridization pattern identical to that of deltapine 90 and sicala (not shown).

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.

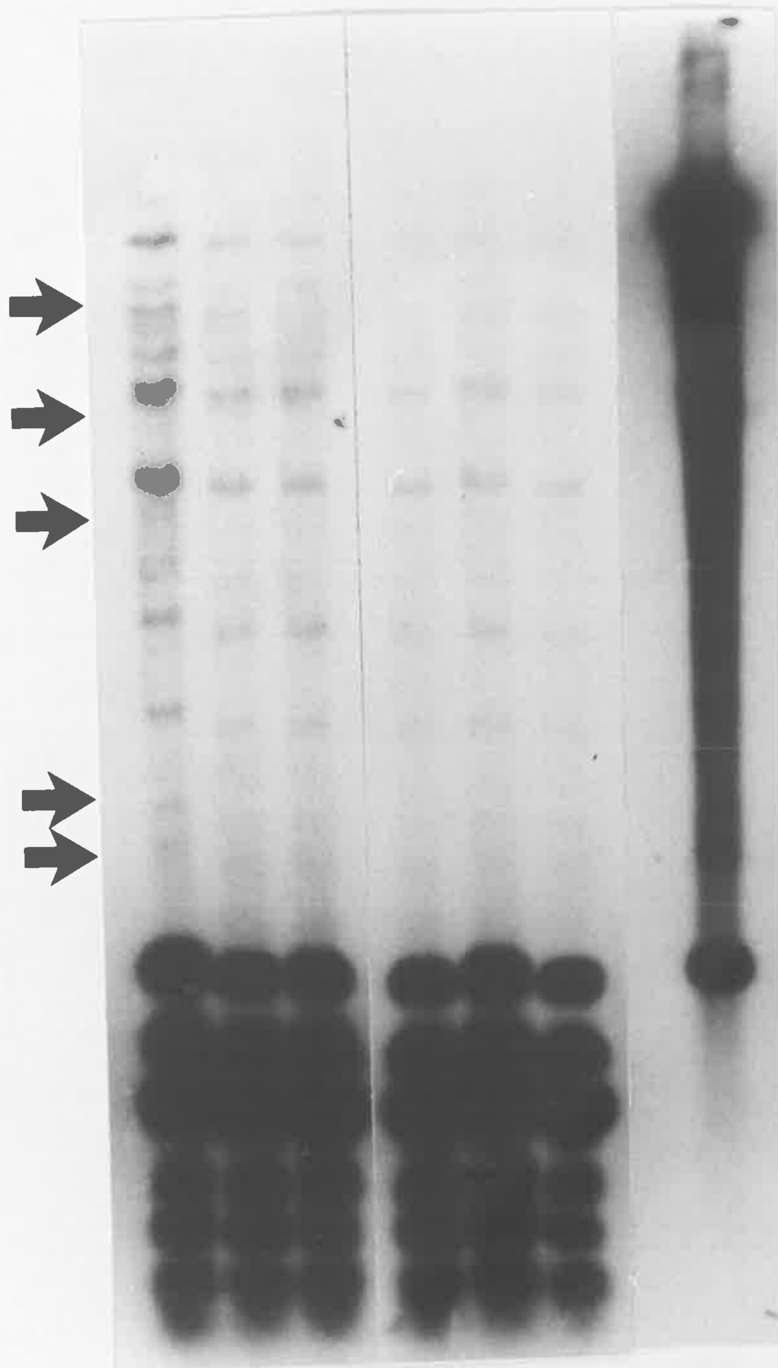

1
2
3
4
5
6
7

Figure 3.3: Homogeneity of nuclear sequences with plastid DNA homology amongst individual *Chenopodium quinoa* plants.

Hybridization of NH1 to total leaf DNA of *C. quinoa*. Each track contained 5 µg of DNA that was isolated from a single plant and restricted with HpaII + EcoRI.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3, 2.0 kbp.

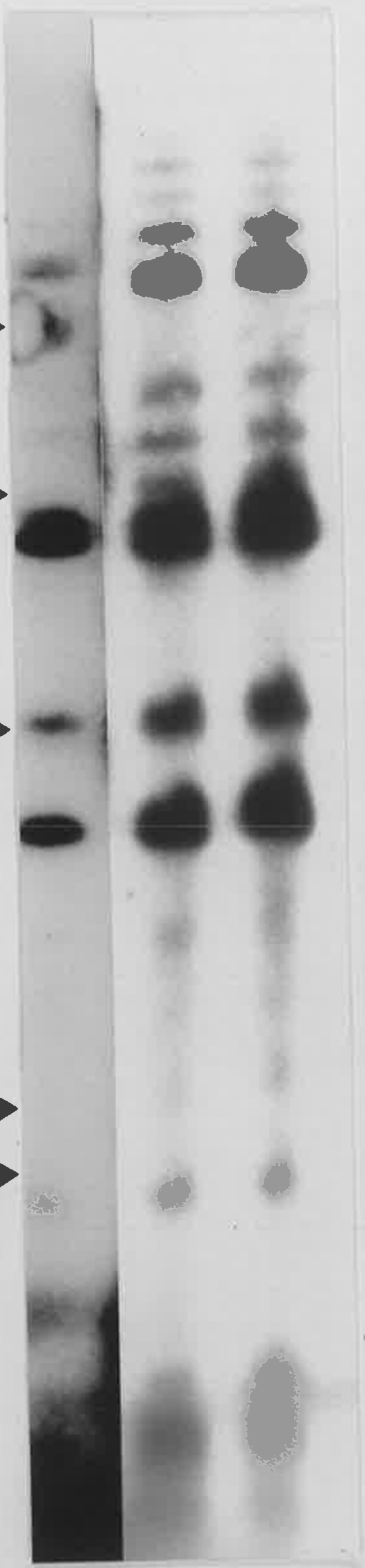

1
2
3

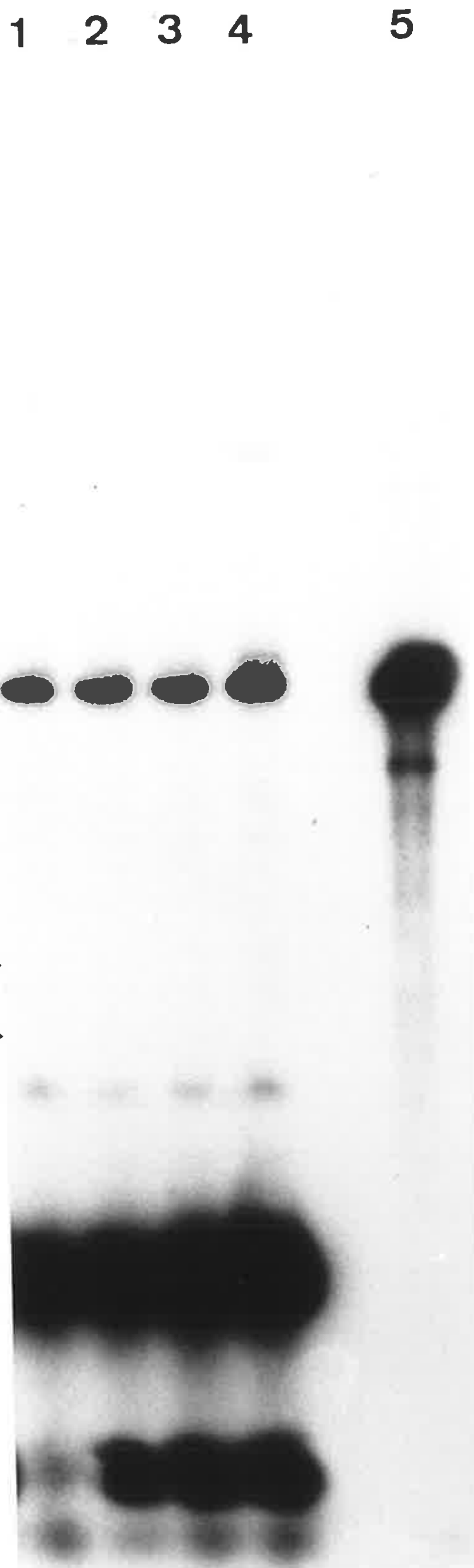
Figure 3.4: A constant hybridization pattern amongst tobacco (*Nicotiana tabacum*) plants, of nuclear sequences with plastid DNA homology.

A) Hybridization of the spinach *rbc*L gene to tobacco total leaf DNA. Each track contained 5 µg of DNA that was isolated from a single tobacco plant and restricted with HpaII + EcoRI, except for lane 5 which contained 0.5 µg of DNA restricted with EcoRI.

B) Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to the same Southern blot membrane used in A.

Arrows indicate molecular weights of 23, 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.

A



B

1 2 3 4 5

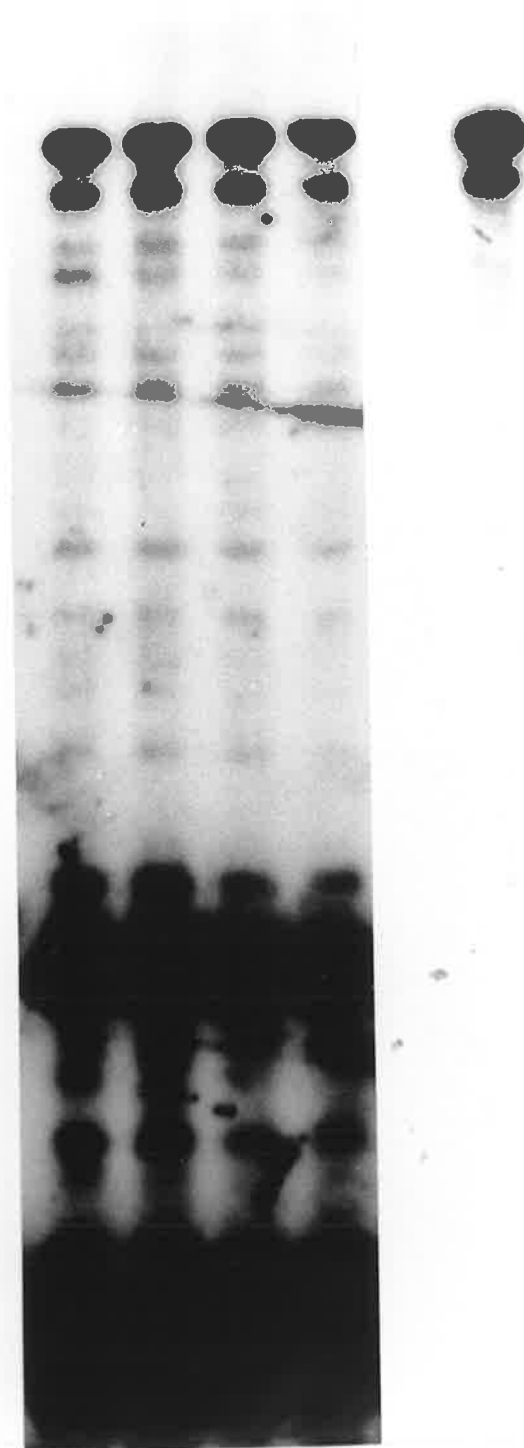


Figure 3.5: A variable hybridization pattern of nuclear sequences with plastid DNA homology is observed amongst different lines of pea (*Pisum sativum*).

Hybridization of NH1 to HpaII + EcoRI restricted total leaf DNA of pea. Each track contains 5 μg of DNA isolated from multiple plants of a single pea line.

Arrows indicate molecular weights 9.4, 6.6 and 4.4 kbp.

1 2 3 4 5 6 7 8

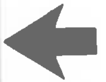
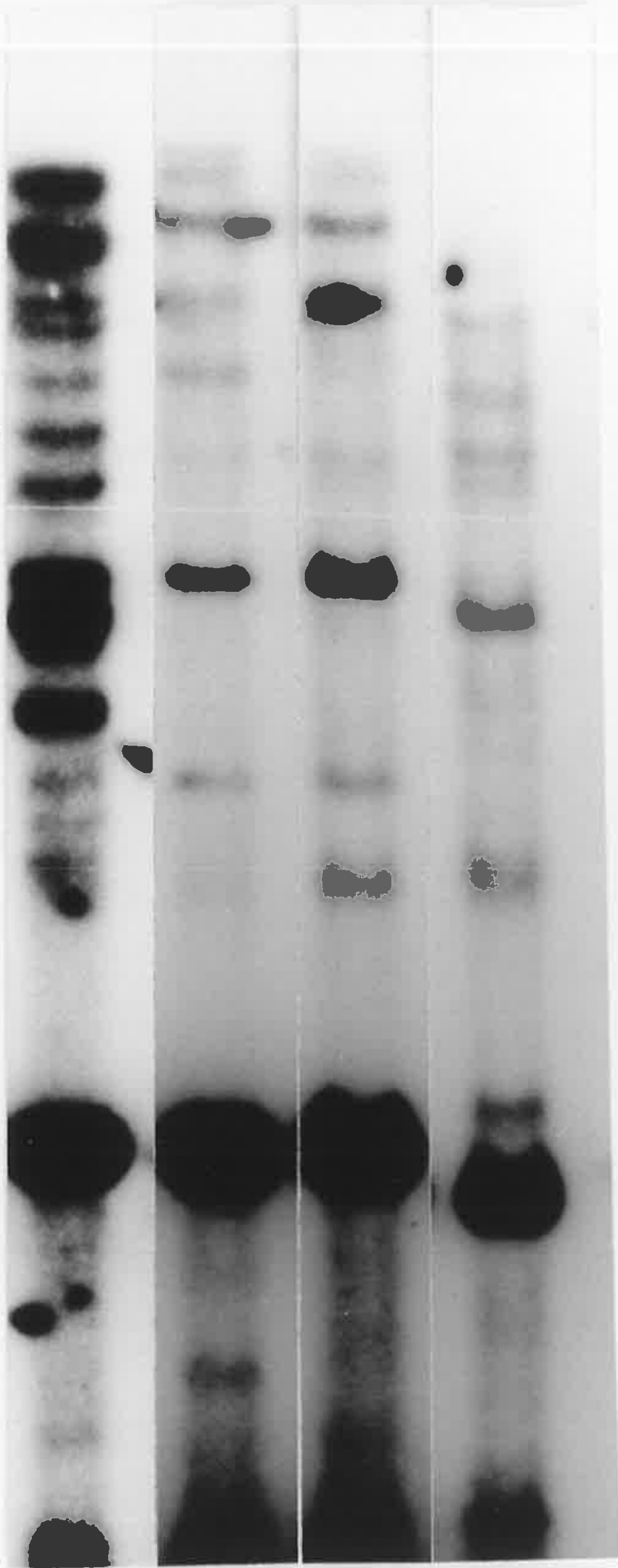


Figure 3.6: A unique spectrum of nuclear sequences with plastid DNA homology is possessed by different varieties of barley (*Hordeum vulgare*).

Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to 5 µg of HpaII + EcoRI restricted total DNA isolated from multiple plants of barley varieties;

1) Swift
2) Cleremont-France
3) Hiproly
4) India 115

Arrows indicate molecular weights of 9.4, 6.6, 4.4, and 2.3 kbp.



1
2
3
4

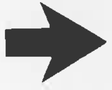
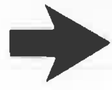



Figure 3.7: Homogeneity within barley varieties of nuclear sequences with plastid DNA homology.

Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to DNA isolated from individual plants of different barley varieties. Each track contains 5 µg of total leaf DNA isolated from a single plant and restricted with HpaII + EcoRI. DNAs in tracks 1-3, 4-5 and 6-8 were isolated from barley varieties Galleon, Hiproly and Swift respectively.

No difference in hybridization pattern was observed amongst three individual plants of the barley variety Arivat (not shown).

Arrows indicate molecular weights of 23, 9.4, 6.6 and 4.4 kbp.

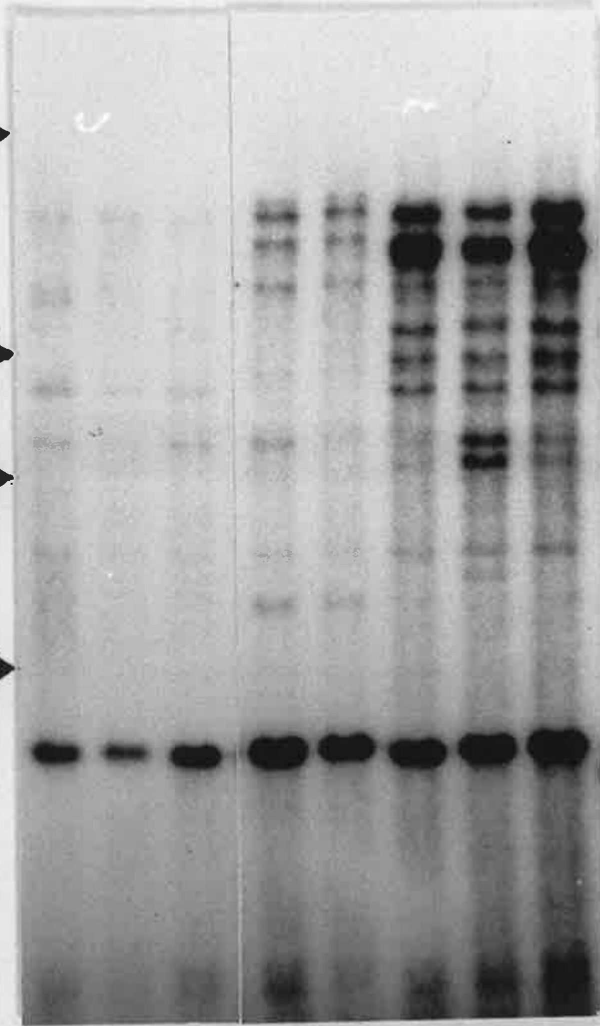

1 2 3 4 5 6 7 8

Figure 3.8: Variation amongst spinach plants of the hybridization pattern of nuclear sequences with plastid DNA homology.

Hybridization of NH1 to HpaII + EcoRI restricted spinach total leaf DNA. Each track contains 5 µg of DNA isolated from an individual spinach plant. Tracks 2, 3, 4 and 6 show an identical spectrum of nuclear homologies, while a 5.5 kbp band is absent in track 1 and an additional 9 kbp band is present in track 5.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3, and 2.0 kbp.

1 2 3 4 5 6

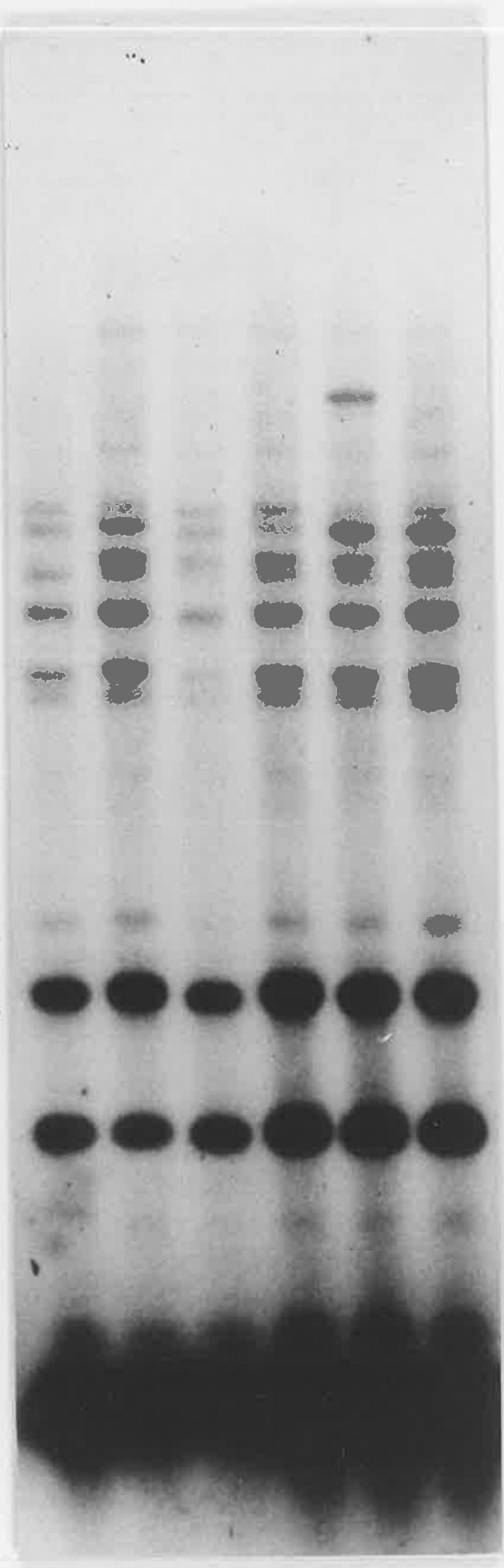


Figure 3.9: Variation amongst *Beta vulgaris* varieties of the hybridization patterns of nuclear sequences with plastid DNA homology.

Hybridization of NH1 to total leaf DNA isolated from *B. vulgaris* var. silverbeet (tracks 1 and 2) and *B. vulgaris* variety cylindrica (tracks 3 and 4). Each lane contains 5 µg of total leaf DNA restricted with either HpaII (tracks 2 and 4) or HpaII + EcoRI (tracks 1 and 3).

A third variety of *B. vulgaris*, crimson globe, also showed a unique hybridization pattern (not shown).

Arrows represent molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.

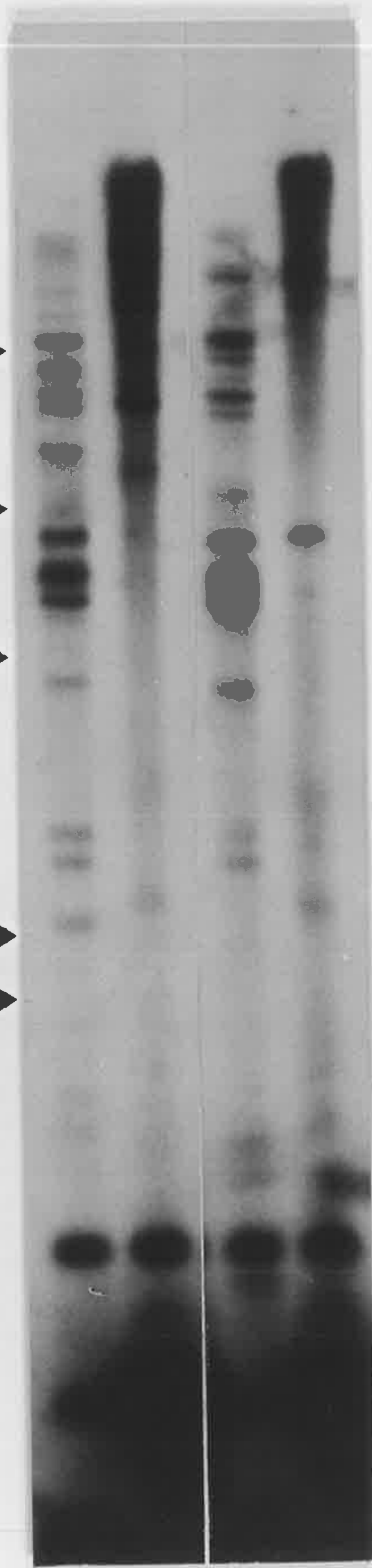

1 2 3 4

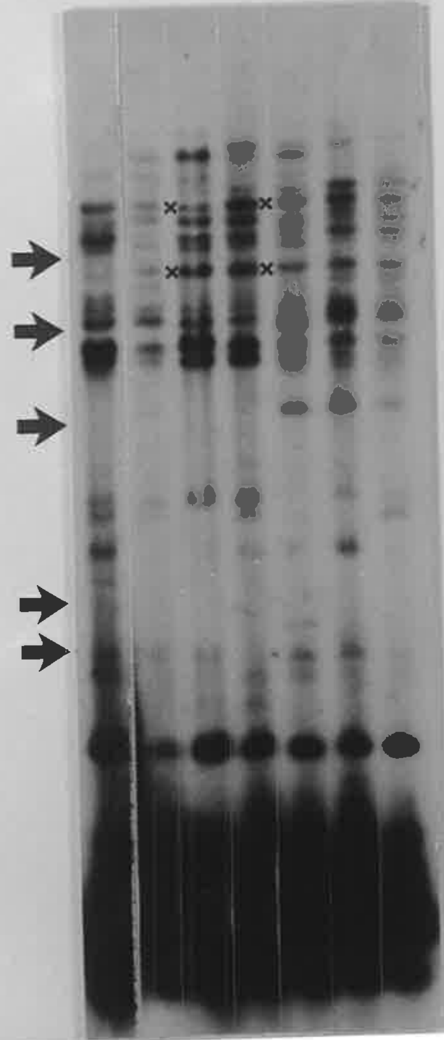
Figure 3.10: A unique spectrum of nuclear sequences with plastid DNA homology is observed for individual plants of *B. vulgaris* var. silverbeet.

A) Hybridization of NH1 to total leaf DNA of *B. vulgaris* var. silverbeet. Each track contained 5 µg of DNA isolated from a single plant and restricted with HpaII + EcoRI. Crosses in tracks 3 and 4 identify bands discussed in the text.

B) Hybridization of the spinach *rbc*L gene to the same Southern blot membrane used in A.

C) Hybridization of pCU18 to the same Southern blot membrane used in A and B.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.


A
1 2 3 4 5 6 7

B
1 2 3 4 5 6 7

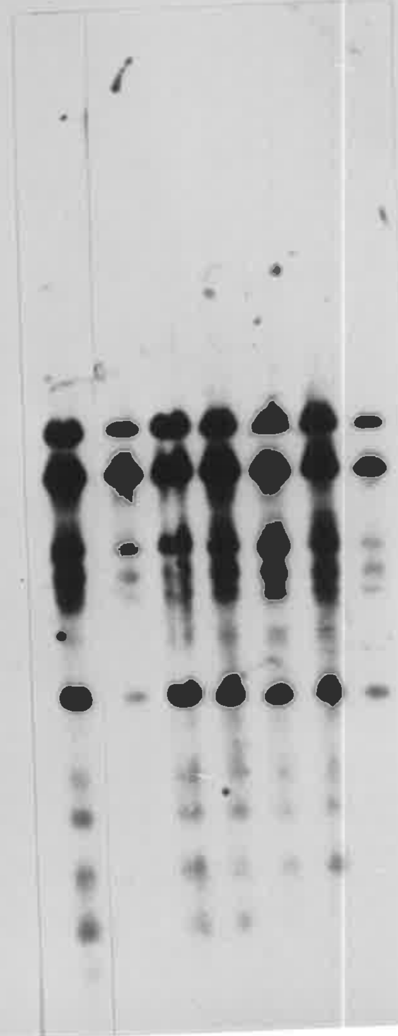

C
1 2 3 4 5 6 7

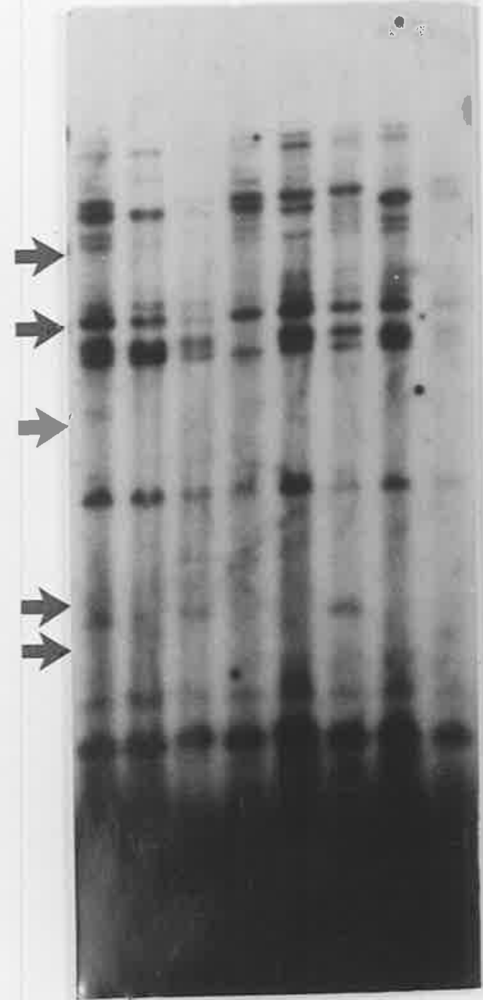
Figure 3.11: A unique spectrum of nuclear sequences with plastid DNA homology is observed for individual plants of *B. vulgaris* var. crimson globe.

A) Hybridization of NH1 to total leaf DNA of *B. vulgaris* var. crimson globe. Each track contained 5 μg of DNA isolated from a single plant and restricted with HpaII + EcoRI.

B) Hybridization of the spinach *rbc*L gene to the same Southern blot membrane used in A.

C) Hybridization of pCU18 to the same Southern blot membrane used in A and B.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.


A
1 2 3 4 5 6 7 8

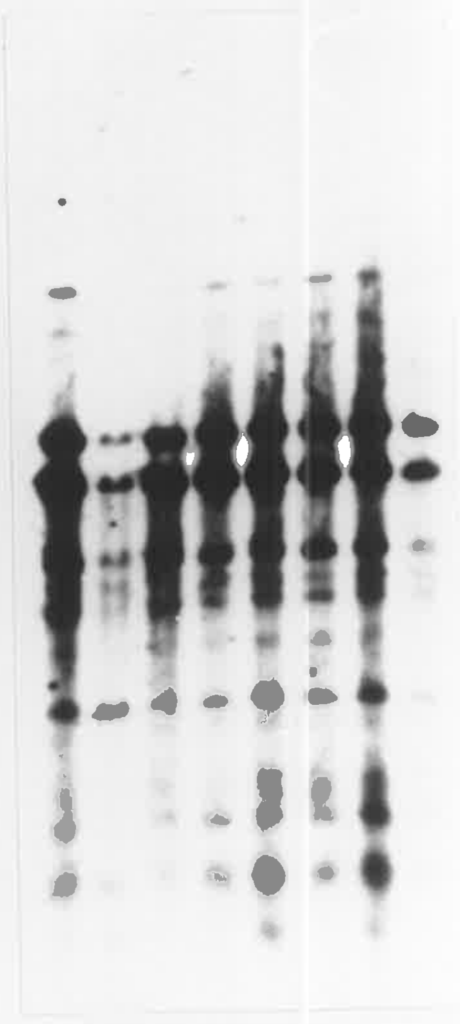
B
1 2 3 4 5 6 7 8

C
1 2 3 4 5 6 7 8

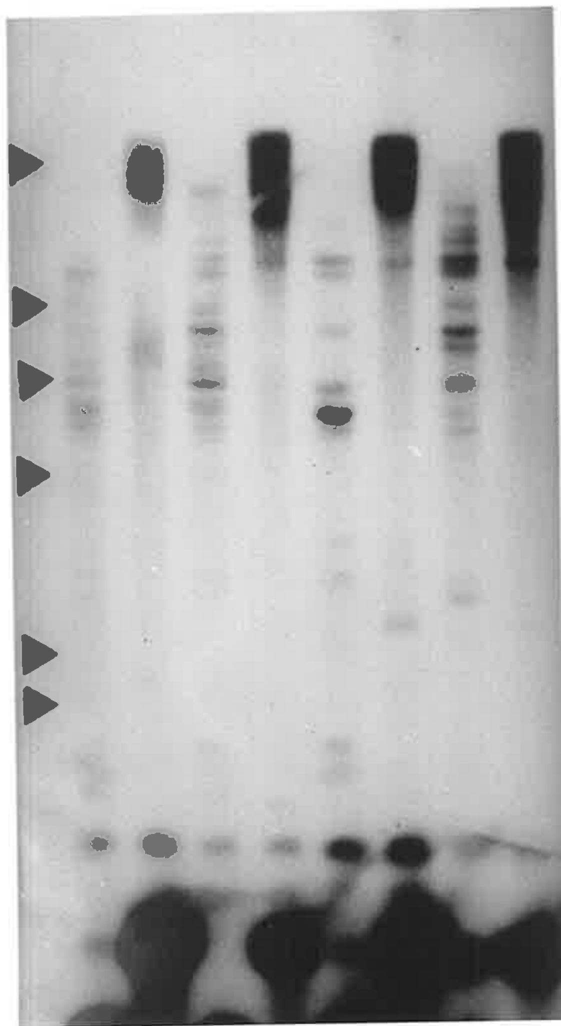
Figure 3.12: Intraplant variation of nuclear sequences with plastid DNA homology in *B. vulgaris* var. silverbeet.

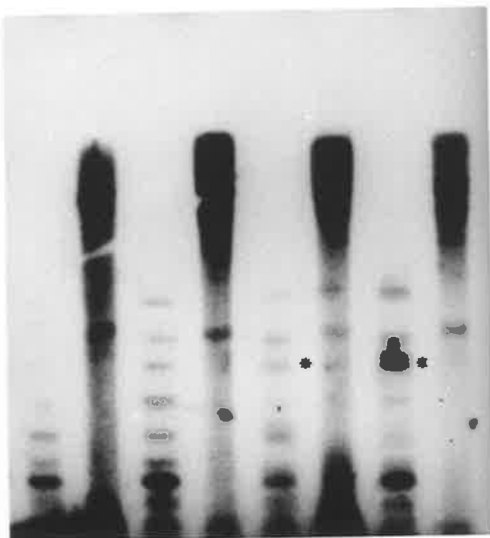
A) Hybridization of NH1 to total leaf DNA of *B. vulgaris* var. silverbeet. Each track contained 5 µg of DNA restricted with either HpaII (tracks 2, 4, 6 and 8) or HpaII + EcoRI (tracks 1, 3, 5 and 7). Tracks 1 and 2 contain DNA isolated from the leaf tissue of a single silverbeet plant while tracks 3 and 4 contain DNA isolated from petiole tissue of the same plant. Tracks 5-6 and 7-8 contain DNA isolated from the leaf and petiole tissue respectively of a second silverbeet plant. A different spectrum of nuclear sequences with plastid DNA homology is observed for both the leaf and petiole tissue of this second plant.

B) Hybridization of the spinach plastid *rbc*L gene to the same Southern membrane used in A. Asterixes identify two bands of approximately 6.6 kbp in lane 7 while only a single 6.6 kbp band is observed in lane 5.

C) Hybridization of pCU18 to the same Southern blot membrane used in A and B. The strongly hybridizing 1.5 kbp band found in HpaII + EcoRI restrictions and the 2 kbp band found in HpaII restrictions represent remnants of the *rbc*L hybridization that could not be removed prior to hybridization of pCU18.

Arrows indicate molecular weights of 23, 9.4 , 6.6, 4.4, 2.3 and 2.0 kbp.


A
1 2 3 4 5 6 7 8


B
1 2 3 4 5 6 7 8

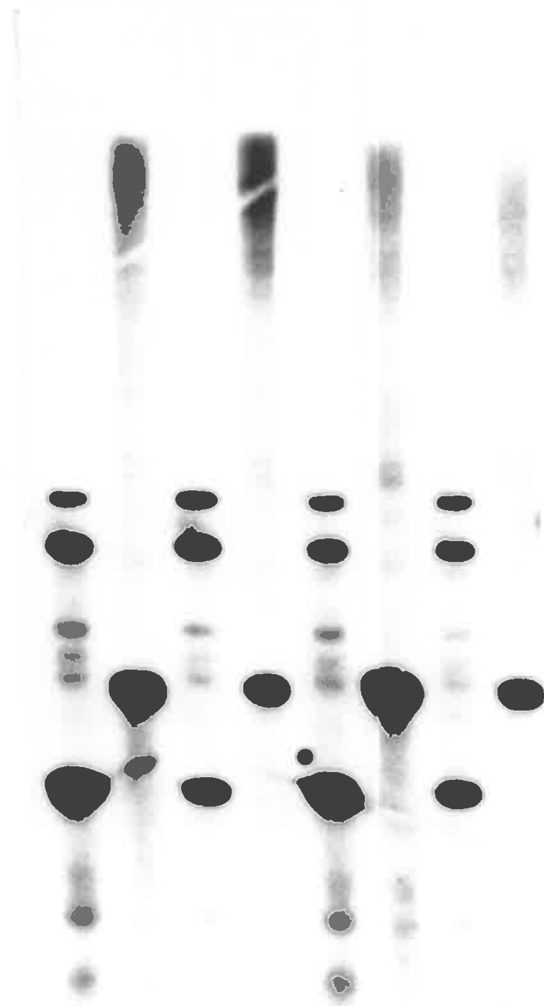
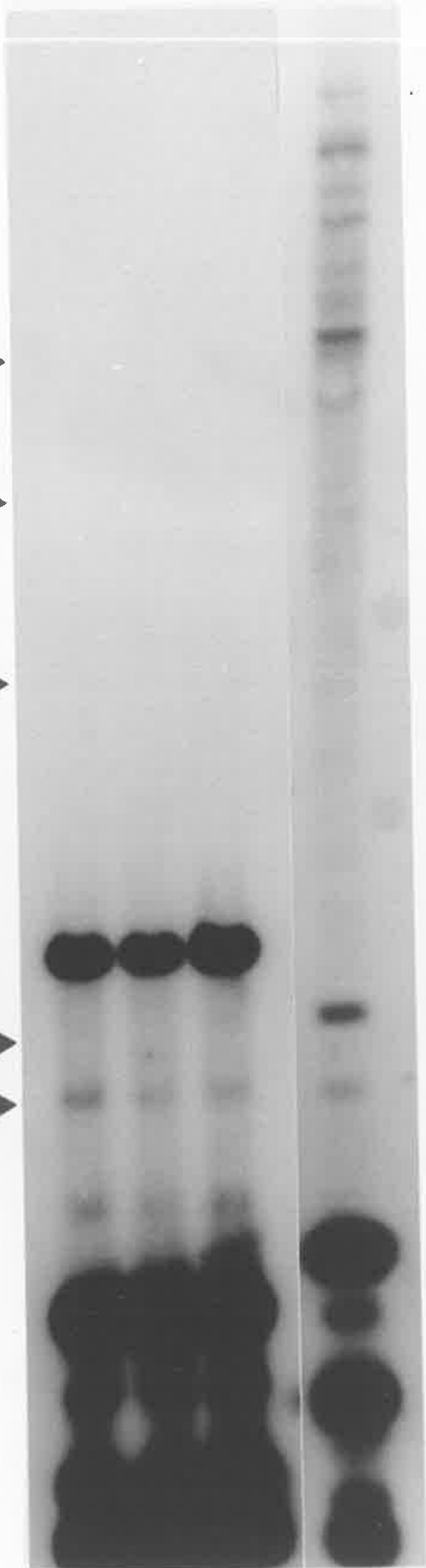

C
1 2 3 4 5 6 7 8

Figure 3.13: Plastid DNA sequence homology can be identified in the nuclear genome of *Tulbaghia violacea* but not in the nuclear genome of *Arabidopsis thaliana*.

Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to total leaf DNA isolated from *A. thalina* (tracks 1-3) and *T. violacea* (track 4). Each track contained 5 µg of DNA isolated from a single plant and restricted with HpaII + EcoRI.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.


1
2
3
4

CHAPTER 4

Promiscuous Plastid Sequences In The Tobacco Nuclear Genome

As the tobacco nuclear genome apparently possesses a relatively high copy number of promiscuous plastid sequences that show conservation of plastid restriction sites, this species was examined in more detail. Hybridization of the spinach *rbc*L gene to tobacco total DNA restricted with EcoRI, revealed a highly homologous 4.4 kbp band (figure 4.1, tracks 3-7) which corresponds to the expected tobacco plastid EcoRI restriction fragment containing the *rbc*L gene (figure 4.2c). When total DNA is further restricted with EcoRI + HpaII a majority of this 4.4 kbp band is reduced to 966 bp, 424 bp and 1700 bp fragments (figure 4.1, track 2). An additional spectrum of bands greater than 2 kbp, with homology to the *rbc*L gene, can also be observed which represent nuclear homologies to this plastid probe. However a significant proportion of the 4.4 kbp EcoRI fragment is not restricted by HpaII and is therefore extensively methylated, suggesting that it is nuclear in origin.

When the EcoRI fragment containing the *rbc*L gene is restricted with KpnI it is reduced to 1018 bp and 3405 bp bands (figure 4.1, track 9; figure 4.2b). The 4.4 kbp EcoRI fragments that are resistant to HpaII digestion are also reduced to bands of 1018 bp and 3405 bp, upon digestion with KpnI (figure 4.1, track 8). Therefore not only do these HpaII resistant fragments comigrate with the *bona fide* plastid EcoRI restriction fragments, but they also possess the same KpnI sites. This suggests that, residing within the tobacco nuclear genome, are large tracts (ie. in excess of 4 kbp) of plastid DNA.

Restriction of total DNA with XhoI, PvuII and SalI shows that the tobacco *rbc*L gene is located on plastid DNA restriction fragments of 8.7 kbp, 6 kbp and 15.5 kbp respectively (figure 4.1, tracks 11, 13, 15). Restriction of total DNA with each of these three enzymes, in conjunction with HpaII, produces a HpaII resistant fragment that comigrates with the *bona fide* plastid DNA fragments homologous to the *rbc*L probe (figure 4.1, tracks 10, 12, 14). These results suggest the presence of even larger tracts of plastid DNA within the tobacco nuclear genome (ie. in excess of 15 kbp). However as

the size of the comigrating HpaII resistant fragment increases its copy number appears to decrease (figure 4.1, compare lanes 2 and 14) and more of the hybridization is observed in high molecular weight unresolved fragments.

Figure 4.2 shows a model which explains the above result. The section of the plastid genome in the vicinity of the *rbc*L gene is bounded by EcoRI sites 4.4 kbp apart and this fragment contains three HpaII sites that would be digested in unmethylated DNA but remain uncleaved if they were modified. The 4.4 kbp fragment therefore often remains in EcoRI + HpaII restricted nuclear DNA because a majority of sequences are sufficently long and conserved to retain EcoRI restriction sites (figure 4.2c). When other restriction enzymes are used in a similar regime the proportion of HpaII resistant DNA falls presumably reflecting either the greater rarity of longer regions of plastid DNA incorporated into the nucleus or the presence of an occasional unmethylated HpaII site. A 15.5 kbp SalI fragment of plastid DNA hybridises the *rbc*L gene probe, and this complete fragment is much more rarely found in methylated DNA than the 4.4 kbp EcoRI fragment (figure 4.1, tracks 2 and 14). However differential methylation sensitivity of these 6-base cutting enzymes could also account for this observation.

Partial digestion is unlikely to explain the presence of these HpaII resistant fragments as each fragment contains multiple HpaII sites (ie. up to 23; figure 4.2b). In a partial digestion at least some of these restriction sites would be expected to be cleaved to give a range of bands characteristic of partial restriction in addition to the undigested fragment. There is no such evidence of partial restriction in these Southern blots.

An analogous result was obtained when the same Southern blot was rehybridized with probes encompassing other regions of the plastid genome. Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment identifies two HpaII resistant EcoRI fragments of 7.3 and 6.7 kbp (figure 4.3, track 2) that comigrate with the *bona fide* plastid EcoRI restriction fragments, homologous to this probe (figure 4.3, tracks, 3-7). An additional spectrum of high molecular weight bands with homology to this plastid DNA probe is also observed. KpnI restriction demonstrates that these two methylated nuclear EcoRI restriction fragments have an identical restriction map to the

bona fide plastid EcoRI restriction fragments (figure 4.3, tracks 8, 9). Again restriction with HpaII plus either XhoI, PvuII or SalI reveals HpaII resistant fragments that comigrate with the plastid restriction fragments homologous to this probe. Hybridization of the tobacco Bam10 fragment (Sugiura *et al*, 1986), which encodes the plastid 16s rRNA gene, to this filter gives a similar result (not shown).

These results suggest that large tracts of plastid DNA (ie. > 15kbp), from several different regions of the plastid genome, are located in the tobacco nuclear genome. The nuclear sequences with homology to the *rbc*L gene are extensively methylated as they are resistant to digestion by both HpaII (figure 4.1, track 2) and MspI (figure 4.1, track 1). Conversely many of the HpaII resistant sequences with homology to the spinach plastid 7.7 kbp PstI DNA fragment (figure 4.3, track 2) are not resistant to MspI restriction (figure 4.3, track 1). Therefore these nuclear sequences with homology to different regions of the plastid genome show differential methylation.

4.1: **Quantification Of Plastid DNA Homologies In Tobacco Nuclear DNA**

To estimate the proportion of total hybridization which could be attributed to these HpaII resistant fragments, standardized loadings of total DNA restricted with EcoRI were included on Southern blots (figure 4.1 and 4.3, tracks 3-7). Scanning of autoradiographs of these tracks with a densitometer enables a curve of DNA amount versus absorbance (ie. relative probe hybridization) to be plotted. By measuring the comparative absorbance of HpaII resistant EcoRI fragments comigrating with plastid EcoRI fragments (figure 4.1 and 4.3, track 2) the proportion of total hybridization signal giving rise to the nuclear homology was determined. An estimated 1.3% of hybridization of total tobacco DNA to the *rbc*L gene is due to the HpaII resistant 4.4 kbp EcoRI fragment present in figure 4.1, track 2. Similarly 0.8% of homology of the hybridization signal to total tobacco DNA by the spinach plastid 7.7 kbp PstI DNA fragment is accounted for by a 7.3 kbp HpaII resistant EcoRI fragment (figure 4.3, track 2) that comigrates with the *bona fide* plastid EcoRI restriction fragment (figure 4.3, tracks 3-7). By determining the number of copies of the plastid genome in this DNA

sample an estimate of the number of large plastid sequences integrated into the nuclear genome can be made.

The polymerase chain reaction (PCR) was used to quantify the number of plastid genomes within this total DNA sample. As PCR yields are affected by minor changes in a number of parameters, direct quantitative comparisions between PCR reactions can not be made. To circumvent this problem an internal standard that can be unambiguously distinguished from the target sequence must be present in each PCR reaction (Sambrook *et al*, 1989). In this case the target sequence was a 450 bp region in the large single copy region of the tobacco plastid genome [from nucleotide 53396 to 53846 of the published sequence (Shinozaki *et al*, 1988)]. Used as an added internal control was a construct molecule consisting of this precise region of the tobacco plastid genome with a 100 bp of pUC19 inserted into the EcoRV site of this plastid sequence (figure 4.4). Therefore both these sequences can be amplified with the same primers and the PCR products resolved by agarose gel electrophoresis. The addition of a known range of this construct molecule to several PCR reactions enables copy number estimates of the target sequence to be made by comparing its PCR yield with that of the internal control.

Figure 4.5 shows a photograph of these PCR products after separation in an agarose gel. From densitometeric scanning of the negative of this photograph and taking into account the different size of these two PCR products a copy number estimate was made for each reaction. Pooling this data gave an estimate of 4350 ± 680 copies of the plastid genome per 1c tobacco nucleus, present in this DNA sample. This value is within the general range observed for other species (Scott and Possingham, 1980).

The number of tobacco nuclear genomes added to PCR reactions was estimated as 128,000 copies based upon genome size measurements obtained from microdensiometry of Feulgen stained nuclei (3.76×10^6 kbp per 1c nucleus, Bennett and Smith, 1976). The final nuclear genome size used to calculate plastid DNA copy numbers was based on dot blots of DNA samples probed with two different tobacco chitinase genes (Neale *et al*, 1990). *N. tabacum* is an allotetraploid with two diploid

ancestors *N. sylvestris* and *N. tomemtosiformis* each contributing a different chitinase gene (Neale *et al*, 1990). This method gave an independant estimate of 100,000 genomes loaded in each PCR reaction, giving a genome size estimate of 4.81 pg per 1c tobacco nucleus. The genome size here is calculated as half the telophase nuclear DNA amount (Bennett and Smith, 1976).

No PCR product was obtained from mitochondrial DNA when using the same primers (chapter 5, figure 5.10) demonstrating that this plastid copy number estimate is not confused by the presence of plastid sequences within the mitochondrial genome. The effect of plastid sequences within the nucleus on this estimate would be negligble as they account for such a small proportion of the total hybridization (0.8 - 1.3%).

From these experiments an estimate of 57 copies of the HpaII resistant 4.4 kbp fragment with homology to the *rbc*L gene (figure 4.1, track 2) and 35 copies of the HpaII resistant 7.3 kbp fragment with homology to the spinach plastid 7.7 kbp PstI DNA fragment (figure 4.3, track 2), per 1c tobacco nucleus is obtained. These values do not include all the sequences within the tobacco nuclear genome which are homologous to these probes, as the fainter spectrum of bands in figure 4.1 and 4.3 (track 2) were not accounted for in this measurement.

Therefore present within the tobacco nuclear genome are many more sequences of plastid origin compared with a spinach haploid nuclear genome where an estimated 5 copies of the plastid genome exist (Steele Scott and Timmis, 1984) and in the tomato nuclear genome where apparently few promiscuous plastid sequences are present (Pichersky *et al*, 1991). In addition some regions of the tobacco plastid genome are more highly represented in the nucleus than others.

4.2: **Nuclear Origin Of These Plastid DNA Homologies**

These results demonstrate that there is a large population of methylated sequences homologous to plastid DNA in the tobacco cell, probably located in the nuclear DNA. There are two other explanations which could account for these relatively high copy number methylated sequences that are homologous to the plastid genome and

possess an apparently identical restriction map. These sequences may be either promiscuous plastid sequences present in the mitochondrial genome or they may represent some form of methylated plastid DNA. Promiscuous plastid sequences have been identified in the mitochondrial genomes of a wide variety of species (Stern and Lonsdale, 1982; Stern and Palmer, 1984). Plastid DNA methylation has been detected in tomato, maize and cultured sycamore cells (Ngernprasirtsiri *et al*, 1988a, 1988b, 1989; Gauly and Kossel, 1989), although refuted by Marano and Carrillo (1991) in tomato. However this methylation generally involves only a few selected sites in the plastid genome and would not account for the extensive methylation observed here (figure 4.2b).

To rule out these other two possiblities a tobacco nuclear DNA preparation was made by differential centrifugation and utilzition of TritonX-100 in extraction buffers. The hybridization experiments were then repeated with the nuclear DNA preparation. The *rbc*L gene again hybridized to a 4.4 kbp EcoRI restriction fragment in both total and nuclear enriched DNA that had been restricted with HpaII + EcoRI, as would be expected if the hybridization was to nuclear fragments (figure 4.6a, tracks 1, 2). Comparision of both nuclear and total DNA restricted with HpaII + EcoRI shows that there is a decrease in plastid DNA in the nuclear preparation, as determined by a decrease in the intensity of hybridization to small plastid HpaII restriction fragments (figure 4.6a, tracks 3, 4). There is however no concomitant decrease in hybridization to the HpaII resistant 4.4 kbp EcoRI fragment present in this nuclear DNA preparation (figure 4.6a, track 1) when compared to total DNA (figure 4.6a, track 2). Therefore this resistant HpaII fragment is of nuclear origin as it copurifies with nuclear DNA. In addition to this 4.4 kbp band there is an identical spectrum of high molecular weight nuclear DNA fragments with homology to the *rbc*L gene probe in both total and nuclear DNA preparations (figure 4.6a, tracks 1, 2).

An analogous result was obtained upon hybridization with the spinach plastid 7.7 kbp PstI DNA fragment. A reduction in hybridization to plastid HpaII fragments is evident in nuclear DNA when compared with total DNA (figure 4.6b, tracks 3, 4).

There is however no obvious reduction in the HpaII resistant fragments (figure 4.6b, tracks 1, 2) that comigrate with the *bona fide* plastid DNA EcoRI restriction fragments (figure 4.6b, track 5, 6). Clearly these sequences are also of nuclear origin. Again an identical spectrum of high molecular weight nuclear fragments with plastid DNA homology, in both total and nuclear DNA is present (figure 4.6b, tracks 1, 2). Marked with an arrow (figure 4.6b, track 2) is a sequence with plastid DNA homology that is substantialy reduced in copy number in the nuclear DNA preparation (figure 4.6b, track 1) and is not expected from the plastid DNA restriction map. This sequence is presumably of mitochondrial origin.

4.3: **Construction Of A Tobacco Genomic Library**

To clone and isolate these plastid sequences in the tobacco nuclear genome it was necessary to produce a nuclear genomic library devoid of cytoplasmic organellar sequences. Total tobacco DNA, isolated from green leaf tissue, was run on a sucrose gradient and the highest molecular weight DNA fraction collected. This DNA was then partially restricted with Sau3A, run on a second sucrose gradient and the resultant 15-20 kbp DNA fraction retained. These fragments consisted of a mixture of nuclear, mitochondrial and plastid sequences. Plastid and mitochondrial sequences were removed from this fraction by extensive restriction with HpaII followed by a third sucrose gradient from which the remaining 15-20 kbp DNA fragments were collected. The remaining DNA fragments clonable in a lambda replacement vector would have been highly enriched for nuclear DNA. These 15-20 kbp fragments were then ligated into the BamHI site of lambda EMBL4 and the ligation packaged using a Packagene^{TM} packaging mix. The resultant library could not be considered as representative of the entire tobacco nuclear genome as hypomethylated nuclear sequences may not have been cloned by the procedure adopted.

4.4: **Plating Of The Genomic Library**

Three restriction systems have been identified in *E. coli* that are specific for methylated DNA sequences (for review see Blumenthal *et al*, 1989). Two of these

re riction systems, termed mcrA and mcrB (for modified cytosine restriction), specifically cleave DNA sequences that contain 5-methylcytosine (Raleigh and Wilson, 1986). Both of these restriction systems have been identified in many *E. coli* strains commonly used as cloning hosts (Raleigh *et al*, 1988; Woodcock *et al*, 1988; 1989). Consequently using *mcr*A$^+$, *mcr*B$^+$ *E. coli* cells as a host in which to clone methylated DNA sequences is likely to result in a dramatic reduction in cloning efficiencies. Plant DNA, which can have up to 40-50% of all cytosine bases methylated, would be expected to be particularly prone to cleavage by these restriction systems (Blumenthal, 1989). For example plating of a *Petunia hybrida* lambda genomic library on a *mcr*B$^-$ host resulted in as much as a 222 fold increase in library titre when compared with an *mcr*B$^+$ host (Woodcock *et al*, 1988).

As the tobacco genomic library produced in this study was made from DNA sequences that were selected on the basis of being methylated it was particularly important to use an appropriate *E. coli* host. Plating of this library onto NW2, a *mcr*A$^-$, *mcr*B$^-$ *E. coli* strain (Woodcock et al, 1989), resulted in a 200 fold increase in library titre when compared to the same library titred on NM538 (*mcr*A$^-$, *mcr*B$^+$) and NM539 (*mcr*A$^-$, *mcr*B$^+$) (Raleigh *et al*, 1988),(table 4.1). To ensure that this difference in library titre was not just the result of differential plating efficiencies of these three *E. coli* strains, wild type lambda EMBL4 was plated using the same plating cells. No significant difference in plating efficiency between NW2 and NM538 could be detected (table 4.2). No plaques were detected when NM539 was used as a phage host because of *Spi* selection (Kaiser and Murray, 1985; Sorge, 1988).

After amplification by passage in *E. coli* strain NW2, an aliquot of the library was plated onto NM538 and NM539 to determine by *Spi* selection what proportion of the library was recombinant (table 4.3) (Sorge, 1988; Kasier and Murray, 1985). Of the 1.2×10^6 pfu contained in this library 79% were recombinant bacteriophage.

Table 4.1

Replicate	NW2	NM538	NM539
1	440	1	2
2	378	0	2
3	472	1	1
Mean pfu	430	1	2

Number of pfu obtained when equivalent dilutions of the tobacco genomic library were plated onto 3 different *E. coli* hosts.

Table 4.2

Replicate	NW2	NM538	NM539
1	2260	1644	0
2	1816	2228	0
3	1864	1988	0
Mean	1980	1953	0

Comparision of plating efficiencies of wild type lambda EMBL4 onto three different *E. coli* strains. Equivalent dilutions of lambda were plated with the above host strains and the number of plaques produced recorded.

Table 4.3

Replicate	NM538	NM539	% Recombinants
1	118	103	
2	132	102	
3	145	108	
Mean	131.7	104.3	79%

After passage in NW2, an equivalent dilution of the genomic library was plated with 3 separate cultures of *E. coli* strains NM538 and NM539. Differential plating efficiencies as a result of *Spi+* selection demonstrate that 79% of the library consists of recombinant bacteriophage.

4.5: Screening Of The Genomic Library

The tobacco genomic library was plated onto agarose plates and screened with three plastid DNA probes; these being the 1.8 kbp EcoRI fragment of pSOCE48 which encodes the spinach *rbc*L gene (Zurawski *et al*, 1981), the 7.7 kbp PstI DNA fragment of the spinach plastid genome (Palmer and Thompson, 1981) and the Bam8 fragment of the tobacco plastid genome (Sugiura *et al*, 1986). Screening of the library with the 7.7 kbp PstI DNA fragment was done at high stringency (Tm -25°C) and identified five positive clones. The library was screened at a reduced stringency (Tm -30) with the other two probes in an attempt to isolate a more divergent population of sequences. This screening identified approximately 300 positive clones (figure 4.7).

Southern hybridization suggested similar numbers of nuclear sequences with homology to the *rbc*L gene (figure 4.1), the 7.7 kbp PstI DNA fragment (figure 4.3) and the Bam8 fragment (not shown). Presumably the large difference in the number of positive clones identified by these two screenings is attributable to the reduced hybridization stringency when using the *rbc*L and Bam8 probes. DNA was isolated from 22 of these 300 plaques and all showed homology to either *rbc*L or Bam8 by Southern hybridization.

Six lambda clones with homology to at least one of these plastid probes were examined in more detail and are described in the following chapters.

4.6: Summary

The nuclear genome of tobacco contains plastid DNA sequence homologies up to at least 15 kbp in size that show conservation of plastid restriction sites, suggesting that large tracts of plastid DNA have integrated into the tobacco nuclear genome. Large plastid sequence homologies include many different regions of the plastid genome (see chapter 5) and are present in estimated copy numbers ranging from 35-57 copies per 1c tobacco nucleus. These sequence homologies can not be explained by a mitochondrial origin or by the possible presence of extensively methylated plastid genomes in the cell as no apparent reduction in copy of these sequences is evident in a nuclear DNA

preparation. Screening of a tobacco nuclear genomic library (0.95 x 10^6 pfu) with three plastid DNA probes covering 9% of the plastid genome yielded some 300 positive clones, consistent with the hypothesis that multiple copies of these sequences are present within the tobacco nuclear genome.

Figure 4.1: Hybridization of the spinach *rbc*L gene to total tobacco DNA.

Each track was loaded with 5 μg of tobacco total DNA except for lanes 3-7 and 9 which were loaded with 0.5, 0.25, 0.125, 0.0625, 0.03125 and 0.5 μg of DNA respectively. DNA samples were digested with the following restriction endonucleases: 1) MspI + EcoRI, 2) HpaII + EcoRI, 3-7) EcoRI, 8) HpaII + EcoRI + KpnI, 9) EcoRI + KpnI, 10) HpaII + XhoI, 11) XhoI, 12) HpaII + PvuII, 13) PvuII, 14) HpaII + SalI, 15) SalI. Markers on the left represent molecular weights of 23, 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.

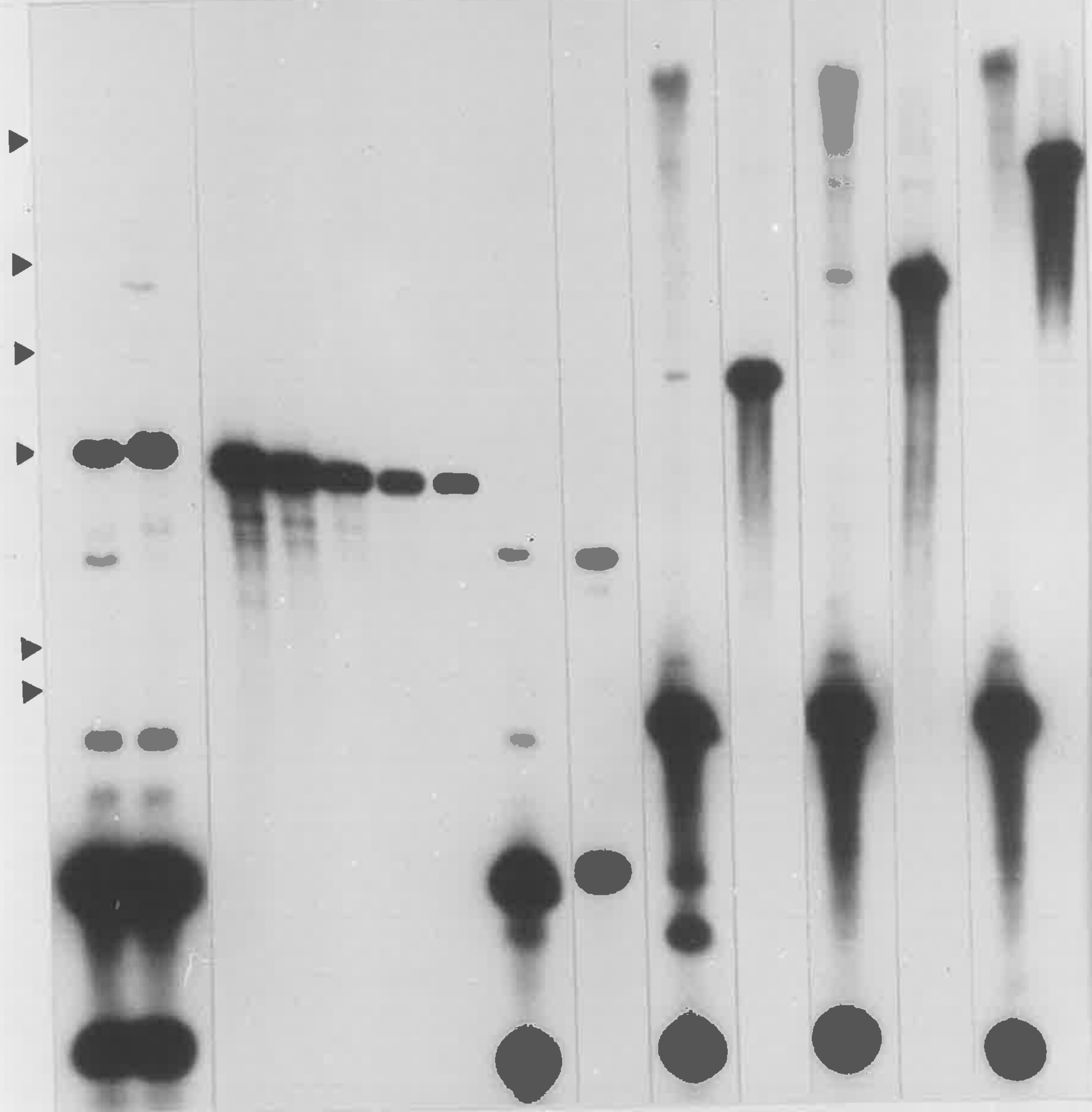
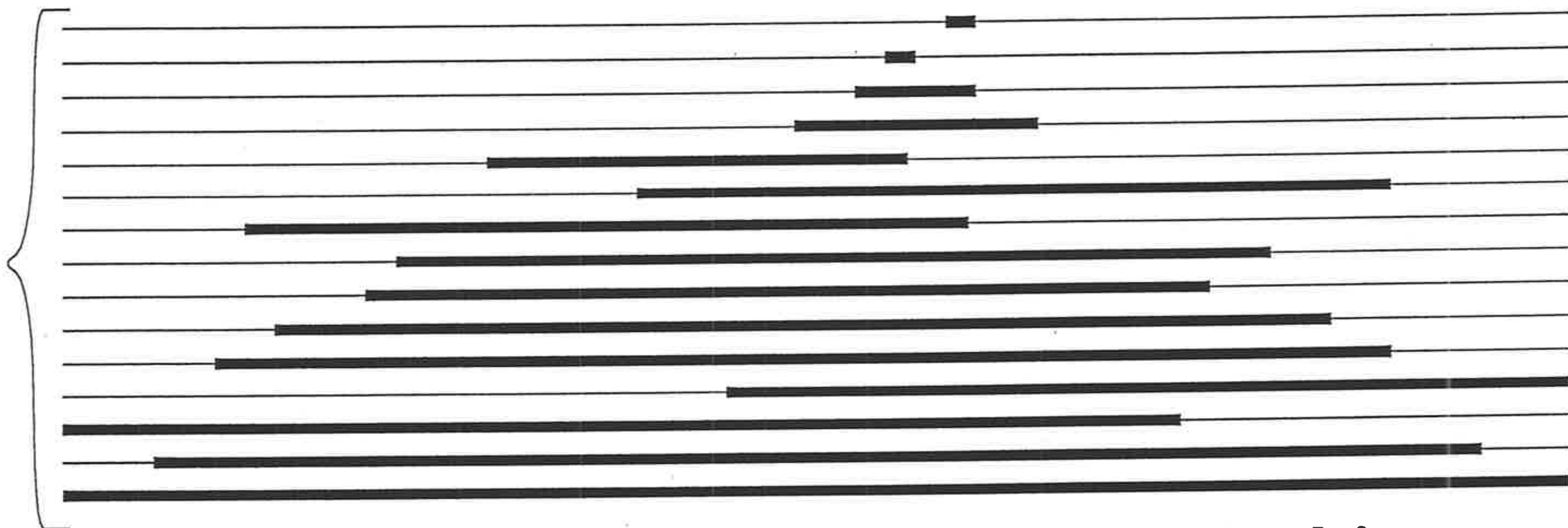

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Figure 4.2: Schematic diagram of hypothetical tobacco plastid DNA fragments inserted into various regions of the nuclear genome.

Nuclear fragments (a) contain at least a proportion of the *rbc*L gene. Promiscuous plastid sequences are represented by a bold line with flanking nuclear sequences represented by a fine line. The size and extent of homology of these sequences to the plastid genome can be determined by a direct comparision to the tobacco plastid *rbc*L gene and flanking regions (b). In (b) letters S,E,X,P,K and H refer respectively to the positions of SalI, EcoRI, XhoI, PvuII, KpnI and HpaII restriction sites in this region of the plastid genome. The position on this restriction map of the 4.4 kbp EcoRI and 15 kbp SalI fragments discussed in relation to figure 4.1 are shown in (c).

2 kb

(a)



(b)

(c)

Figure 4.3: Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to total tobacco DNA.

Each track was loaded with 5 µg of DNA except for lanes 3-7 and 9 which were loaded with 0.5, 0.25, 0.125, 0.0625, 0.03125 and 0.5 µg of DNA respectively. DNA samples were restricted with the following restriction endonucleases: 1) MspI + EcoRI, 2) HpaII + EcoRI, 3-7) EcoRI, 8) HpaII + EcoRI + KpnI, 9) EcoRI + KpnI, 10) HpaII + XhoI, 11) XhoI, 12) HpaII + PvuII, 13) PvuII, 14) HpaII + SalI, 15) SalI. Markers on the left represent molecular weights of 23, 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.

The strongly hybridizing band of 4.4 kbp found in tracks 1-7 represent remnants of the *rbc*L gene hybridization (figure 4.1) that could not be removed from the Southern membrane.

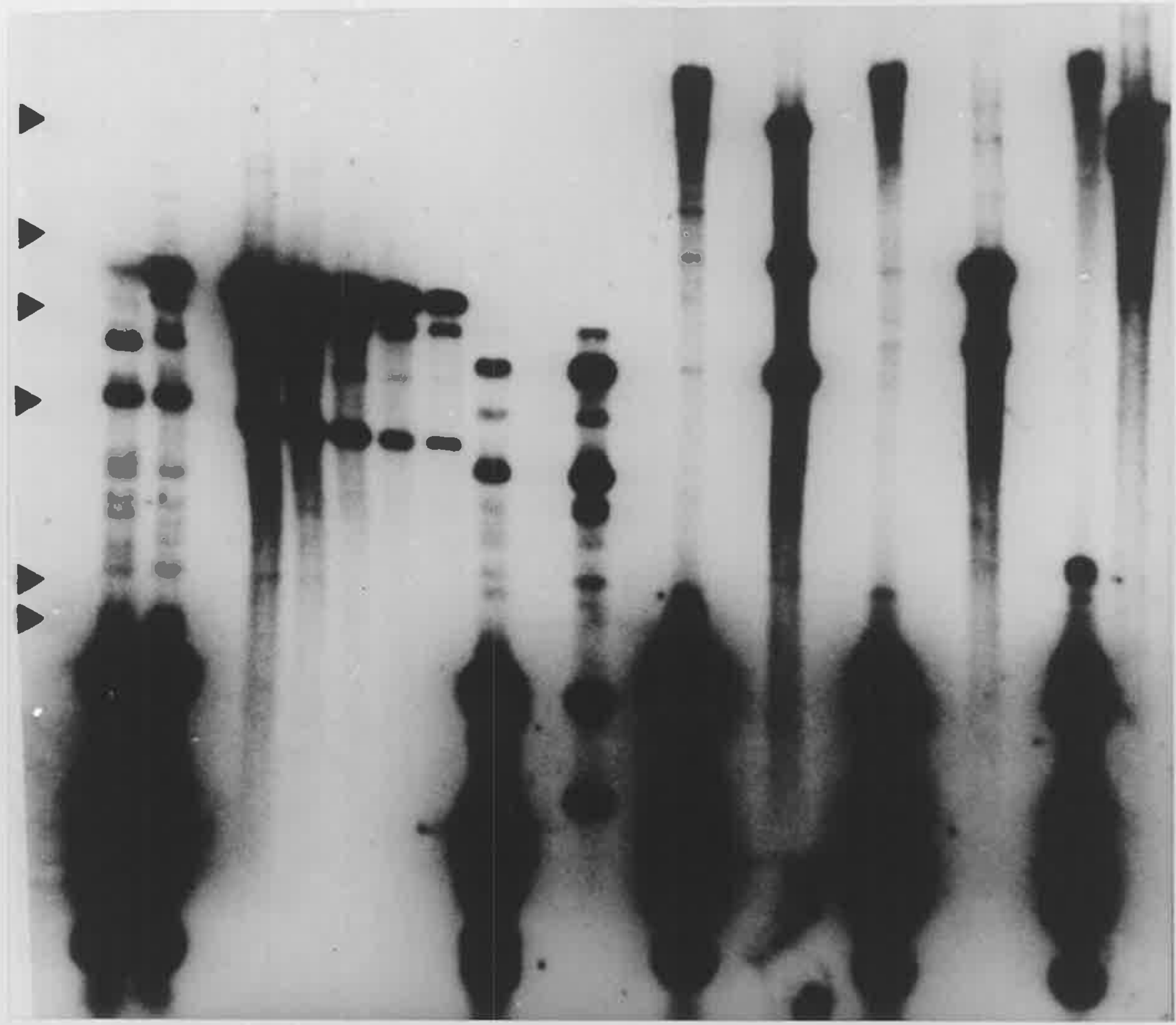
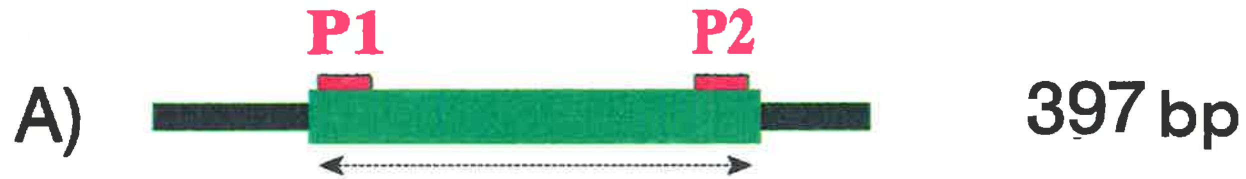

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

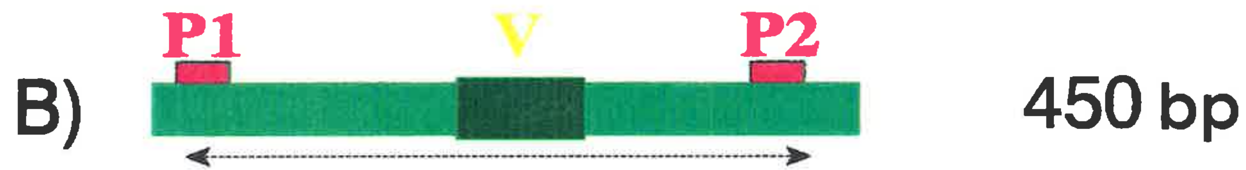
Figure 4.4: Tobacco DNA sequences amplified by PCR.

PCR primers (P1 and P2) were designed to amplify a 450 bp region of the tobacco plastid genome from nucleotide 53396 to 53846 (b). The sequence of these primers is presented in chapter 5, figure 5.8. Added to each PCR reaction was a construct plasmid consisting of a 1 kbp BglII fragment of pTB19 (which included nucleotides 53,396 to 53,846 of the tobacco plastid genome (Shinozaki *et al*, 1986), inserted into the BamHI site of pUC19. Into the EcoRV (yellow V) site of this cloned plastid sequence (b) was inserted an additional 100 bp HaeIII fragment of pUC19, shown in purple in (c). PCR amplification with primers P1 and P2 therefore amplifies a 550 bp sequence from this plasmid construct. The region of the tobacco nuclear sequence contained in lambda clone 3D, that has a 12 and 41 bp deletion and is discussed in chapter 5, is represented in (a). PCR amplification of this molecule using primers P1 and P2 results in the production of a 397 bp sequence.

cpDNA insert in nDNA



unrestricted true cp DNA



construct

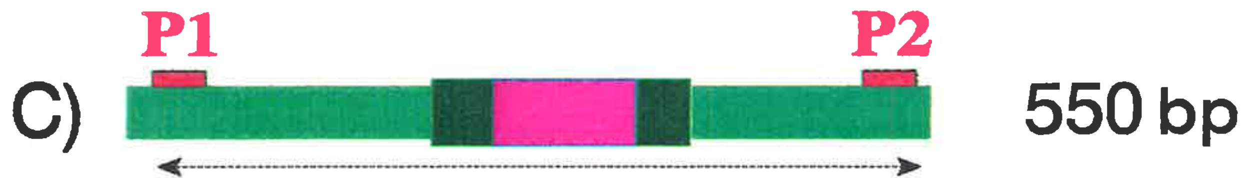


Figure 4.5: **PCR quantification of the number of plastid genomes contained in total tobacco DNA.**

2% agarose gel showing the PCR products obtained from PCR reactions containing primers P1 and P2 (Figure 4.4 and chapter 5, figure 5.8) and:

1) 0.5 µg of total tobacco DNA plus 10 copies of plasmid construct added per 1c tobacco genome.
2) 0.5 µg of total tobacco DNA plus 1000 copies of plasmid construct added per 1c tobacco genome.
3) 0.5 µg of total tobacco DNA plus 1500 copies of plasmid construct added per 1c tobacco genome.
4) 0.5 µg of total tobacco DNA plus 2000 copies of plasmid construct added per 1c tobacco genome.
5) 0.5 µg of total tobacco DNA plus 2500 copies of plasmid construct added per 1c tobacco genome.
6) 0.5 µg of total tobacco DNA plus 3000 copies of plasmid construct added per 1c tobacco genome.
7) 0.5 µg of total tobacco DNA plus 3500 copies of plasmid construct added per 1c tobacco genome.
8) 0.5 µg of total tobacco DNA plus 15,000 copies of plasmid construct added per 1c tobacco genome.
9) 0.5 µg of tobacco total DNA.
10) No DNA added to the PCR reaction.

Track 11 contained 1 µg of wild type lambda DNA restricted with HindIII. The smallest band visible in this track is 564 bp in size.

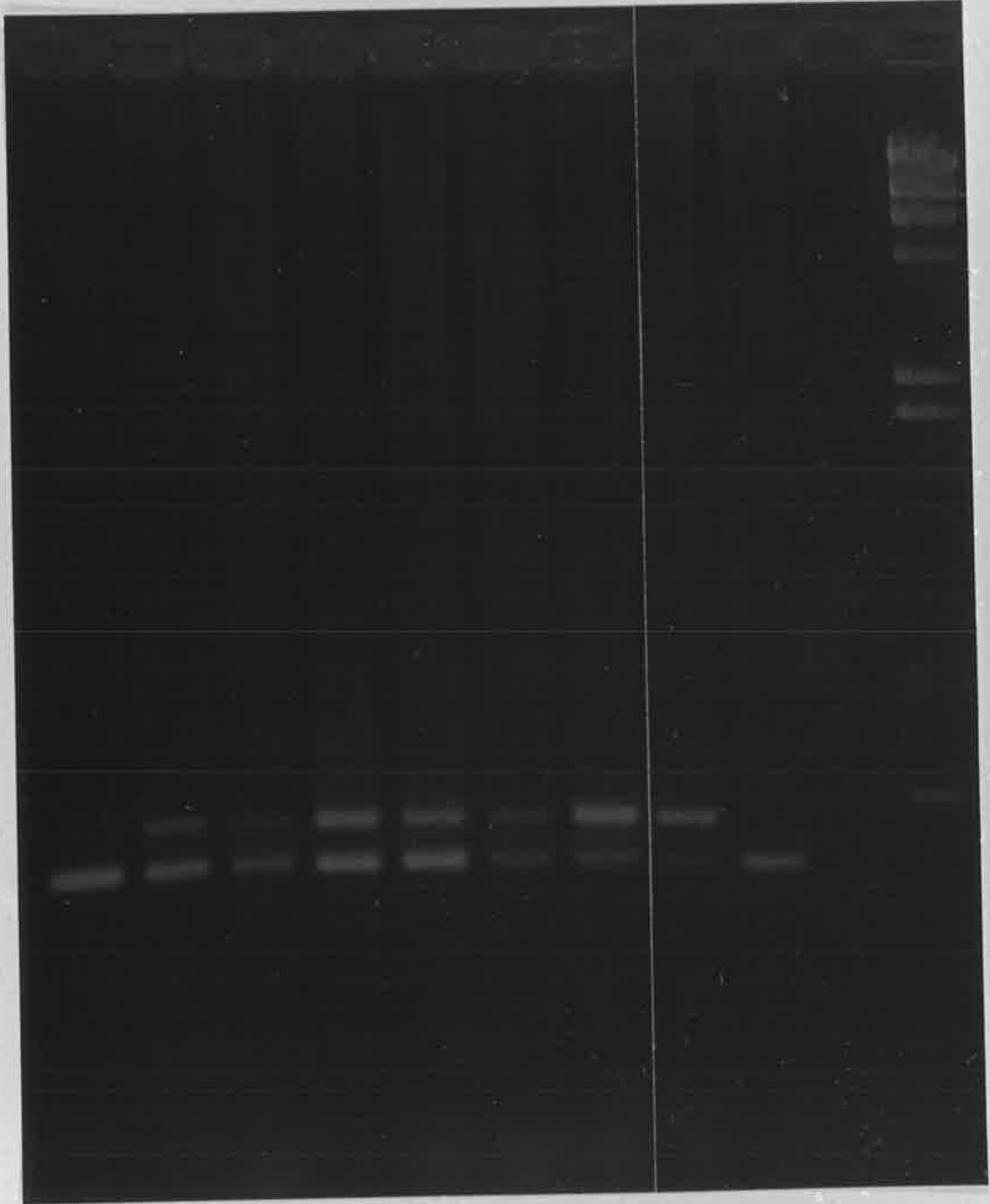
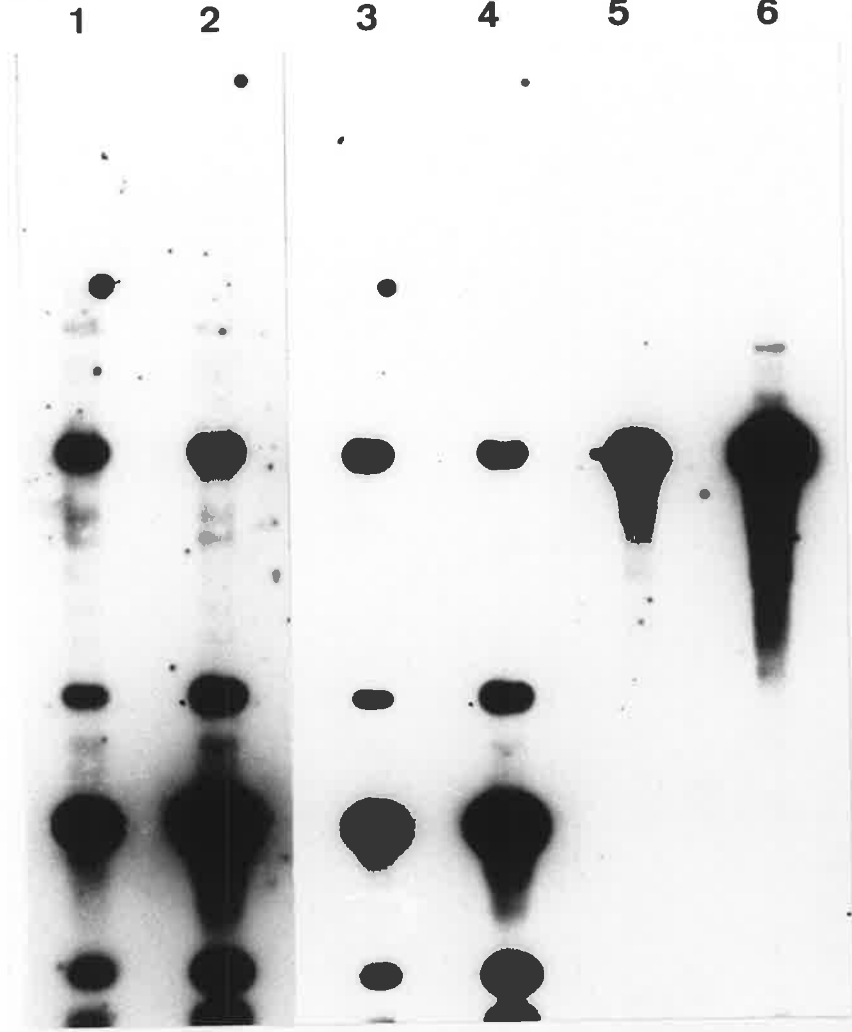

1 2 3 4 5 6 7 8 9 10 11

Figure 4.6: Hybridization of plastid DNA probes to tobacco nuclear and total DNA.

A) Hybridization of the spinach plastid *rbc*L gene to tobacco nuclear DNA (lanes 1, 3 and 5) and total DNA (lanes 2, 4 and 6). Lanes 1-4 were loaded with 5 µg of DNA that was restricted with HpaII + EcoRI while lanes 5 and 6 contained 1 µg of DNA restricted with EcoRI. Lanes 1 and 2 are a longer exposure of lanes 3 and 4.

B) Rehybridization of the nitrocellulose filter used in A) to the spinach plastid 7.7 kbp PstI DNA fragment. The curved arrow indicates a homology putatively assigned to mitochondrial DNA.

Arrows indicate molecular weights of 23, 9.4, 6.6, 4.4, 2.3 and 2.0 kbp.


A
1
2
3
4
5
6

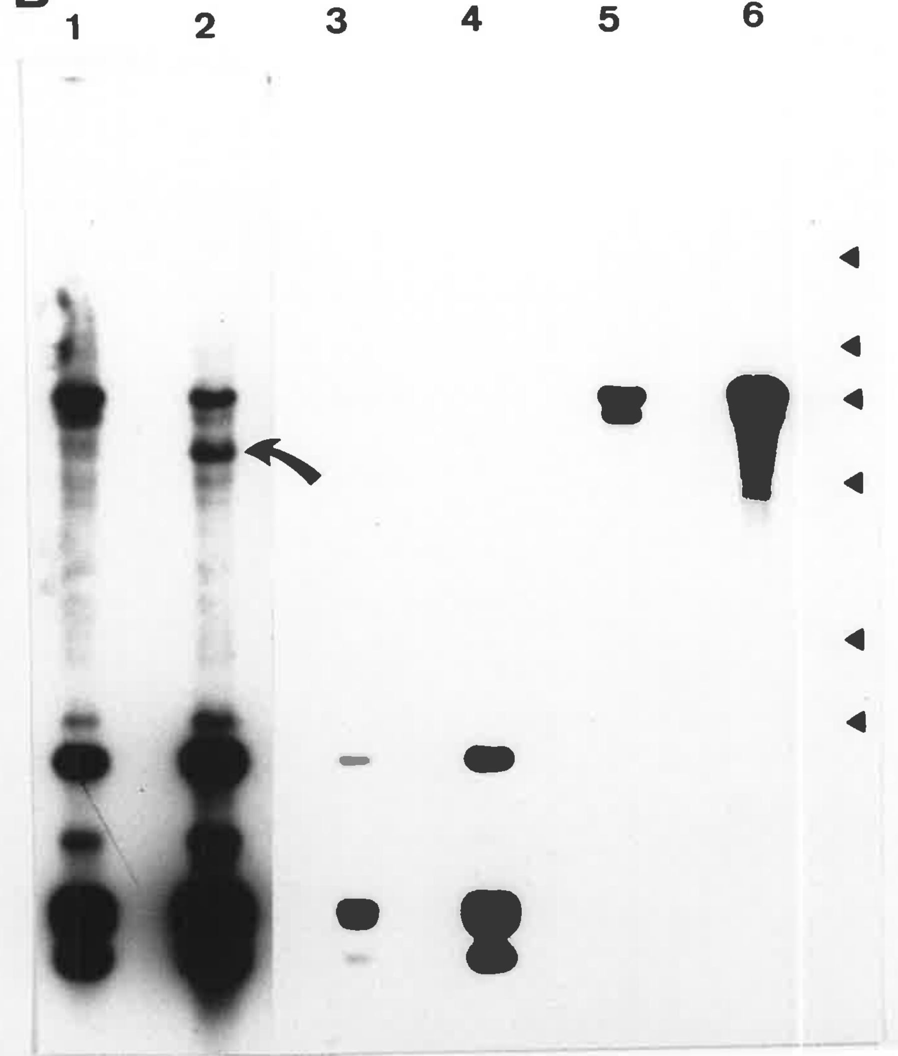
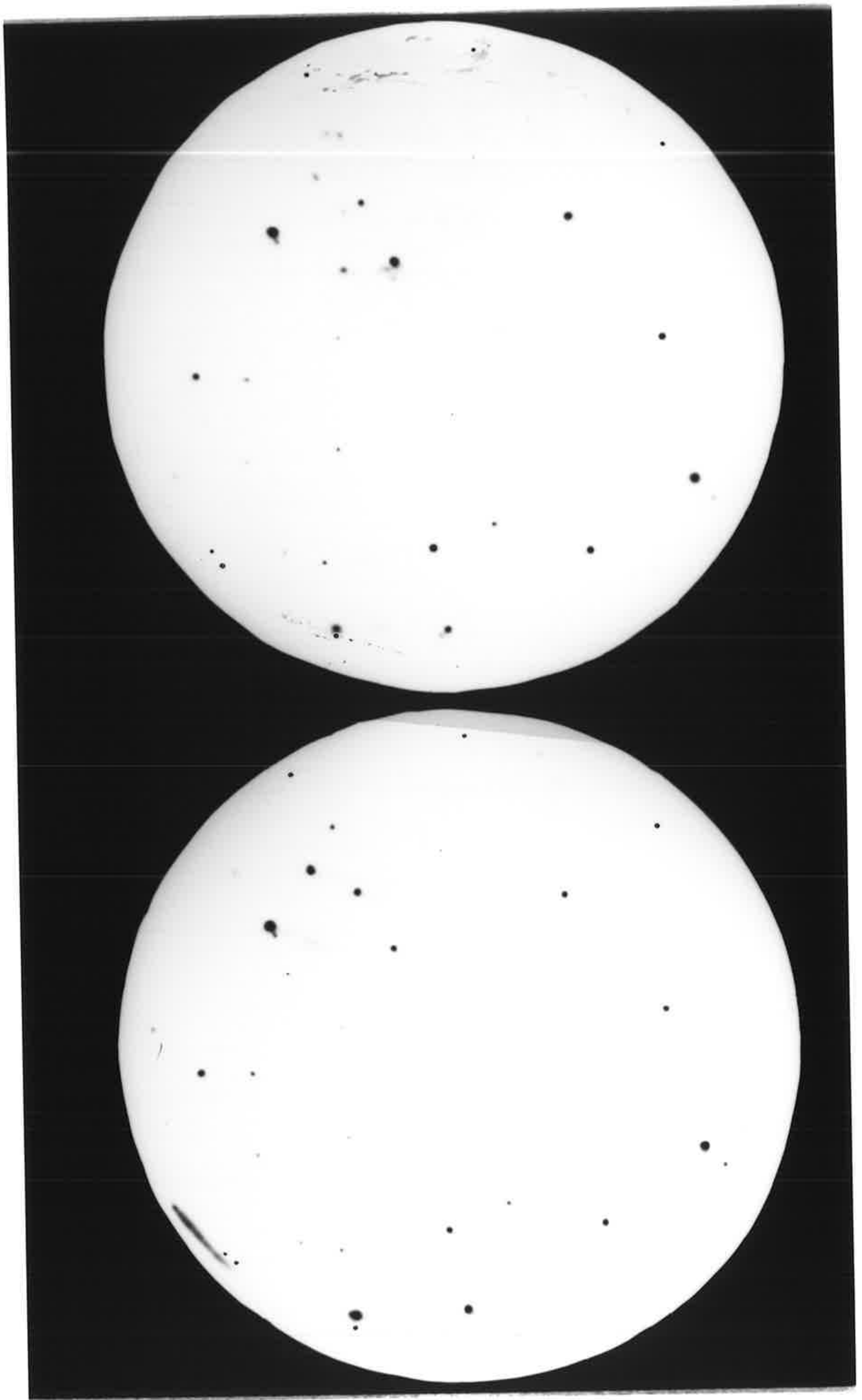

B
1
2
3
4
5
6

Figure 4.7: Plaque lifts.

Hybridization of the spinach *rbc*L gene and the tobacco plastid Bam8 DNA fragment to duplicate plaque lifts of the tobacco nuclear genomic library.



CHAPTER 5

Characterization of lambda clones 3D, 3b, 3-2, 1H and 2d

Six lambda clones isolated from the tobacco genomic library were examined more closely, in varying degrees of detail. Three of these clones (called 3D, 3b and 3-2) were isolated by screening the genomic library with the spinach plastid 7.7 kbp PstI DNA fragment, while the remaining three clones (1M, 1H, 2d) were identified by screening the library at reduced stringency with a mixture of the spinach plastid *rbc*L gene (Zurawski *et al*, 1981) and the tobacco plastid Bam8 fragment (Sugiura *et al*, 1986). A reduced hybridization stringency was adopted for these latter two probes in an attempt to isolate a more diverse population of promiscuous plastid sequences.

Hybridization of total tobacco mitochondrial DNA to EcoRI restricted DNA from five of these lambda clones identified regions of the clones homologous with the mitochondrial genome (figure 5.1). However every lambda clone possessed at least one EcoRI restriction fragment that was not homologous to the tobacco mitochondrial genome, thereby precluding a mitochondrial origin for each clone. Of the EcoRI restriction fragments that were homologous to the mitochondrial genome some did not have a co-migrating analogue in the mitochondrial DNA preparation (figure 5.1b, track labelled "MT"), providing additional evidence that these clones are not mitochondrial in origin. However all the lambda clones possessed two EcoRI restriction fragments that consisted of both vector and cloned DNA sequences and these would not be expected to have a co-migrating mitochondrial DNA analogue. The homology between these lambda clones and the mitochondrial genome is the result of homology between plastid derived sequences present in both the mitochondrial and nuclear genomes.

Of the six lambda clones examined, five were homologous in their entirety to the tobacco plastid genome while the remaining clone (1M) consisted of contiguous nuclear and plastid-derived DNA sequences. The clones completely homologous to the

tobacco plastid genome (clones 3D, 3b, 2D, 3-2, 1H) are described in this chapter while clone 1M is the subject of chapter 6.

5.1: **Lambda Clone 3D**

Clone 3D was isolated from the tobacco genomic library by virtue of its homology with the spinach plastid 7.7 kbp PstI DNA fragment and found to contain a 13.8 kbp DNA insert. Homology between this plastid probe and clone 3D was confined to a single 3.8 kbp EcoRI restriction fragment (figure 5.2a, track 1). Subsequent hybridization of pTB19 [which encompasses the tobacco plastid genome from nucleotide 43426 to 55921 (Sugiura *et al*, 1986)] to clone 3D showed homology between every EcoRI restriction fragment of this clone and the plastid probe (figure 5.2b). The intensity of hybridization of these EcoRI restriction fragments appeared equivalent to their stoichiometry, suggesting that the cloned DNA was completely homologous to the plastid genome. In addition six EcoRI restriction fragments were common to both clone 3D and pTB19 (figure 5.2a) further confirming that this clone contained a large tract of plastid-like DNA.

Six out of the nine EcoRI restriction fragments that constituted clone 3D (marked with arrows in figure 5.2b, track 1) were subcloned into pUC19 and either partially or completely sequenced. All these sequences were highly homologous to the published sequence of the tobacco plastid genome between nucleotide 42455 to 56290 (Shinozaki *et al*, 1986). Each of the remaining three EcoRI restriction fragments of clone 3D comigrated with an EcoRI restriction fragment of pTB19 and showed comparable hybridization intensity to its analogue when probed with pTB19 (figure 5.2b).

From these data clone 3D could be unequivocally assigned as completely homologous to the tobacco plastid genome from nucleotide 42455 to 56290 (figure 5.3), apart from the possibility of minor insertions. The EcoRI restriction map of this clone is very similar to that of tobacco plastid DNA except that a 2.9 kbp fragment contains an additional EcoRI site (boxed in figure 5.3) to give rise to a 1.6 kbp and 1.3 kbp product.

This clone therefore encompasses nuclear equivalents of the chloroplast genes for *psa*A (part), *trn*S, *rps*4, *trn*T, *trn*L, *trn*F, *psb*G, *ndh*C, *trn*V, *trn*M and part of *atp*E in addition to five unidentified open reading frames. Homology between clone 3D and the spinach plastid 7.7 kbp PstI DNA fragment was confined to the first 1.5 kbp of DNA adjacent to the long arm of the vector of this clone (figure 5.3).

Although highly homologous to the tobacco plastid genome, significant sequence differences were identified between this clone and the organelle genome. Four regions of clone 3D were sequenced in both directions as indicated in figure 5.3. Part (2978 bp) of the largest 3.8 kbp EcoRI restriction fragment (figure 5.3, region 1) differed from the published plastid sequence by 17 base substitutions, three deletions and four insertions (figure 5.4). The deletion of a single nucleotide in clone 3D (equivalent to nucleotide 43152 of the tobacco plastid genome) destroys the reading frame of the *psa*A gene in 3D resulting in the occurrence of a translation termination codon 26 bases downstream of this deletion (figure 5.5). The *psa*A gene product is localized in the thylakoid membrane and is an essential component of photosystem I. Consequently as clone 3D does not encode a functional *psa*A gene product, it is unlikely to represent a *bona fide* plastid DNA sequence.

It could be argued that the mutation in this gene may be circumvented by RNA editing. To preserve the *psa*A ORF of clone 3D, three base substitutions and two nucleotide insertions would be required. This would represent an unprecedented amount of editing of a plastid gene and nucleotide insertions have never been detected in the plastid RNA editing process (Hoch *et al*, 1991; Cattaneo, 1992). In addition clone 3b possesses a deletion in the *psa*A gene (see section 5.2) different from that in 3D and so the plant would require two different editing pathways for the same gene.

Clone 3D is clearly not of mitochondrial origin (figure 5.1) and must therefore represent a large tract of plastid DNA that has integrated into the tobacco nuclear genome and subsequently diverged in sequence. This result is consistent with Southern hybridization data which also indicated the presence of large tracts of this region of the plastid genome within the tobacco nucleus (chapter 4).

Region 2 (figure 5.6) of clone 3D is a 328 bp EcoRI fragment which is homologous to an intergenic region of the plastid genome from nucleotide 50844 to 51171. This region differs from the plastid genome by only four base substitutions.

Region 3 of clone 3D is a 528 bp EcoRI fragment which is homologous to the published sequence of the tobacco plastid genome from nucleotide 52233 to 52760 (figure 5.7). This region of the plastid genome contains the *ndh*C gene which encodes a subunit of the plastid NADH dehydrogenase complex. Region 3 differs from the plastid genome by only two base substitutions (figure 5.7), one of which occurs in the *ndh*C open reading frame. However this base substitution does not alter the amino acid sequence of the protein product.

Conspicuous sequence differences occur in region 4 of clone 3D (figure 5.3), which shows homology to the plastid genome from nucleotide 53401 to 53797 and contains two relatively large deletions of 12 and 41 base pairs in addition to seven base substitutions (figure 5.8). Each of these deletions is flanked by an 11 bp direct repeat in chloroplast DNA (figure 5.8) which may have been causal to the deletions. Homologous pairing and recombination between similar direct repeats has been postulated to account for a large deletion in a promiscuous sequence in the rice mitochondrial genome (Moon *et al*, 1988). Alternatively the sequence differences between clone 3D and the plastid genome could have arisen from insertions within the plastid genome subsequent to the transposition of the 3D sequence to the tobacco nucleus.

The deletions observed in clone 3D may have occurred as an artifact of cloning and this possibility was tested by searching for the deletions directly in genomic DNA. To determine whether a sequence identical to clone 3D is present in tobacco nuclear DNA, PCR primers containing EcoRI restriction sites were designed to amplify a 397 bp region spanning the 12 and 41 bp deletions (figure 5.8: chapter 4, figure 4.4). These primers were identical to the primers used to estimate the number of copies of the plastid genome present in a tobacco total DNA preparation, described in chapter 4. The 41 bp deletion which is absent in putative nuclear DNA is present in plastid DNA where it contains an EcoRV restriction site (figure 5.8; chapter 4, figure 4.4). A partially

purified nuclear DNA preparation was therefore restricted with EcoRV to cleave any contaminating plastid DNA present in this sample. Exponential amplification of the 397 bp region was expected to be possible only in sequences lacking the EcoRV site. If EcoRV did not digest all the plastid DNA molecules or nuclear sequences were present that lacked both the deletions and an EcoRV site a 450 bp product would also be exponentially amplified. Gel analysis indicated a single PCR product from EcoRV digested nuclear DNA of the expected length consistent with the presence of a tobacco genomic sequence identical to clone 3D (figure 5.9a). This PCR product was cloned (figure 5.9b, track 1) and its sequence was found to be identical to that of clone 3D (figure 5.8). When total leaf DNA, undigested by EcoRV, was amplified using the same primers a 450 bp product was obtained as predicted from the tobacco plastid DNA sequence (Shinozaki *et al*, 1986). This product was cloned and three inserts were shown to have the published (Shinozaki *et al*, 1986) plastid DNA sequence (figure 5.8). When the same PCR protocol was applied to purified mitochondrial DNA no products were synthesized indicating that this region is not found in the mitochondrial genome (figure 5.10)

PCR was used to quantify the number of plastid derived sequences in the tobacco nuclear genome containing this 41 bp deletion. The protocol adopted was identical to that used to estimate the total number of plastid genomes found in a tobacco total DNA preparation (chapter 4) except that tobacco nuclear DNA that had been restricted with EcoRV was used. The same internal PCR standard was used (Chapter 4, figure 4.4) which consisted of a 100 bp of pUC19 inserted into the EcoRV site of one of the cloned 450 bp PCR products obtained from plastid DNA amplification. Relative to each tobacco nuclear genome, a range of copy numbers of this construct was added to EcoRV digested nuclear DNA and the samples amplified with the same primers. Gel analysis of the products (figure 5.9b) indicated that between 10 and 20 copies of the genomic sequence were present in each 1c tobacco nuclear genome.

This PCR experiment was repeated using a narrower range of added construct sequences and the products labelled by the addition of ^{32}P dATP in the final two

rounds of PCR. These PCR products were then resolved on both agarose and denaturing polyacrylamide gels which enabled estimates to be made that were based on both ethidium bromide staining and densitometry of autoradiographs. Taking into account the relative size and adenine composition of the PCR products this region of clone 3D was estimated to be present 15 ± 5 times in each 1c tobacco nuclear genome. Estimates of the number of nuclear genomes added to each PCR reaction were based upon the same calculations as described in chapter 4

The 4231 bp of clone 3D that were bidirectionally sequenced differed from the published sequence of the tobacco plastid genome (Shinozaki *et al*, 1986) by 30 base substitutions, five deletions and four insertions. Given this high degree of sequence conservation, clone 3D probably represents a sequence very recently transferred from plastid DNA to the nuclear genome. Assuming a rate of sequence divergence similar to that of mammalian pseudogenes [5×10^{-9} changes per site per year (Li *et al*, 1985; Dennis *et al*, 1988)] the 3D sequence is estimated to have transferred to the nucleus approximately 700,000 years ago.

5.2: **Lambda Clone 3b**

Lambda clone 3b was shown to have a restriction map identical to that of the tobacco plastid genome from nucleotide 24962 to 42791 (Shinozaki *et al*, 1986) (figure 5.11). Hybridization of pTB7 and pTB20 (Sugiura *et al*, 1986) identified homology to every restriction fragment of the clone, suggesting that it is also entirely plastid in origin (result not shown)

Regions of this clone that were sequenced all showed very high sequence homology to the tobacco plastid genome from nucleotide 24962 to 42791 (figure 5.12). From the combined results of restriction mapping and sequencing it was concluded that clone 3b contained an 18 kbp DNA insert that was homologous in its entirety to the tobacco plastid genome (apart from minor base substitutions, insertions and deletions). This clone therefore encompasses nuclear equivalents of the following plastid genes: *rpo*B (part), *trn*C, *trn*D, *trn*Y, *trn*E, *trn*T, *psb*D, *psb*C, *trn*S, *trn*G, *trn*fM, *rps*14, *psa*B

and part of the *psa*A gene, in addition to one unidentified open reading frame (ORF 62) (figure 5.12).

Like clone 3D, regions of clone 3b that were sequenced showed very high homology to the tobacco plastid genome (Shinozaki *et al*, 1986). However sequence divergence was present that inactivated an ORF in an essential photosynthetic gene and thereby precluded this clone from representing a *bona fide* plastid sequence. As this clone is not mitochondrial in origin (figure 5.1) it must therefore represent a large tract of nuclear DNA with homology to the plastid genome.

Region 1 (figure 5.12) of clone 3b is a 365 bp EcoRI / SphI restriction fragment that was bidirectionally sequenced. It differed from the tobacco plastid genome by a deletion of a single nucleotide pair at position 42681 (Shinozaki *et al*, 1986) (figure 5.13). This region encompasses the 5' end of the *psa*A gene and the single nucleotide deletion causes the formation of a translation termination codon 36 bp further downstream (figure 5.14). Both of the other potential reading frames of this region contain termination codons (figure 5.14) and therefore further sequence modifications upstream of this region would not reproduce a translatable mRNA.

Region 2 of clone 3b is a 606 bp EcoRI fragment that is homologous to the tobacco plastid genome (Shinozaki *et al*, 1986) from nucleotide 31660 to 32267, but differs by two base substitutions and the deletion of a single nucleotide pair (figure 5.15). This deletion occurs at the equivalent position of nucleotide 32259 of the tobacco plastid genome (Shinozaki *et al*, 1986), which is part of the coding region of the plastid *trn*E gene. Conversely region 3 of clone 3b, which is 1032 bp in size, was identical in sequence to the plastid genome from nucleotide 24961 to 25992 (Shinozaki *et al*, 1986).

The high sequence homology between clone 3b and the tobacco plastid genome suggests that, like clone 3D, clone 3b represents a very recent transfer of plastid DNA to the nuclear genome.

5.3: **Lambda clone 3-2**

Clone 3-2 possessed a restriction map (figure 5.17) that was identical to that of the tobacco plastid genome from approximately nucleotide 28760 to 46300 (Shinozaki *et al*, 1986) (figure 5.16a and b). All restriction fragments of this clone were homologous to either pTB19 or pTB20 (Sugiura *et al*, 1986) suggesting that clone 3-2 was entirely plastid in origin (figure 5.16c). Sequencing of a 605 bp EcoRI restriction fragment of clone 3-2 (figure 5.17) identified extensive sequence homology to the tobacco plastid genome from nucleotide 31660 to 32267 (Shinozaki *et al*, 1986) (figure 5.18).

Sequence divergence was evident, with the 605 bp fragment of 3-2 differing from the plastid genome by two base substitutions and two separate deletions of a single nucleotide pair (figure 5.18). Both of these deletions occurred in tRNA genes and given the conservative nature of plastid tRNA molecules this strongly suggests that clone 3-2 is not a *bona fide* plastid DNA sequence (Shimada *et al,* 1991). The *trn*D, *trn*Y and *trn*E genes of the tobacco plastid genome are 100%, 99% and 96% homologous to their rice counterparts respectively (Shimada *et al*, 1991). This 605 bp EcoRI fragment is the same region of sequence as the 606 bp EcoRI fragment of clone 3b, which did not show homology to the mitochondrial genome (figure 5.1). Therefore clone 3-2 is also not mitochondrial in origin.

From these results it can be concluded that clone 3-2 represents a 17.5 kbp tract of the tobacco nuclear genome that is homologous in its entirety to plastid DNA. This clone therefore encodes nuclear equivalents of the following plastid genes: *trn*C, *trn*D, *trn*Y, *trn*E, *trn*T, *psb*D, *psb*C, *trn*S, *trn*G, *trn*fM, *rps*14, *psa*B, *psa*A in addition to one unidentified ORF (ORF62).

5.4: **Regions of sequence homology between lambda clones**

Lambda clones 3D and 3b both showed sequence homology to the tobacco plastid genome from nucleotide 42456 to 42792. Sequence divergence was evident when this region of the clones was compared (figure 5.19). In this 336 bp overlap each

clone possessed a different deletion of a single nucleotide pair (figure 5.19). Each deletion was responsible for the inactivation of the *psa*A gene in both of these clones.

Sequence comparisons between the 606 bp EcoRI fragment of clone 3b and the 605 bp EcoRI fragment of clone 3-2 identified further sequence divergence (figure 5.20). Clone 3-2 contained 2 single base deletions relative to clone 3b, while clone 3b contained a single base deletion relative to clone 3-2 (figure 5.20). Of interest is the observation that both clone 3b and 3-2 contain a base substitution of a G for an A at the equivalent of nucleotide position 32068 of the plastid genome (Shinozaki *et al*, 1986), and the reverse base substitution at position 32193 (figure 5.15 and 18). Given the unlikelihood of two identical base substitutions following two translocations these two tracts may have arisen by the duplication of a single plastid sequence present within the nuclear genome. Subsequently these two sequences have diverged to account for the differences observed in this region for these two clones. Alternatively this observation could be the result of a sequence polymorphism existing between the plastid genome of the tobacco strain used in this study and the plastid genome sequenced by Shinozaki *et al* (1986). The sequence of this region has been reported twice (Shinozaki *et al*, 1986: Ohme *et al*, 1985). Shinozaki *et al* (1986) report nucleotide 32193 of the tobacco plastid genome as an A (figure 5.15 and 18), while Ohme *et al* (1985) reports this equivalent nucleotide as a G (figure 5.21). Both clone 3-2 and 3b contain a G at this position (figure 5.15 and 5.18). Conversely nucleotide 31680 and 31681 of the tobacco plastid genome (Shinozaki *et al*, 1986) have been deleted in the sequence published by Ohme *et al* (1985) (figure 5.21). Both of these nucleotides were present in the sequence of clone 3-2 and 3b (figure 5.15 and 5.18). As both of these sequence differences are present in clones 3-2 and 3b neither are likely to represent sequencing errors by either group and therefore indicate that polymorphisms occur within the plastid genome of *N. tabacum*. Thus two plastid genomes of tobacco differed by a deletion of two nucleotide pairs and one base substitution in a total of 607 bp, which is a surprisingly large amount of variation given the reported conservation of the plastid genome. Some of the differences between the published sequence of the tobacco plastid genome (Shinozaki *et*

al, 1986) and the lambda clones in this study, may therefore not represent sequence divergence of these nuclear integrants but rather reflect plastid sequence polymorphisms prior or subsequent to the transposition event. However the inactivation of ORFs in essential plastid genes demonstrates unequivocally that these cloned sequences do not represent *bona fide* plastid DNA.

5.5: **Lambda clones 1H and 2d**

Restriction enzyme mapping (figure 5.22a), Southern hybridization (figure 5.22b and 5.22c) and partial sequence analysis (figure 5.23) of lambda clone 1H suggested that this clone was also homologous in its entirety to the plastid genome. Clone 1H encompassed the plastid genome from approximately nucleotide 75000 to 92700 (Shinozaki *et al*, 1986) and therefore encoded nuclear equivalents of the following plastid genes: *psb*B, *psb*F, *pet*B, *pet*D, *rpo*A, *rps*11, *sec*X, *rps*8 *rps*14, *rpl*16 *rps*3, *rpl*22, *rps*19, *rp*12, *rpl*23, *trn*I (figure 5.23). From limited restriction enzyme mapping and sequence analysis clone 2d was tentatively identified as being homologous to the plastid genome from approximately nucleotide 75400 to 89000 (Shinozaki *et al*, 1986).

Neither lambda clones 1H nor 2d have been unequivocally distinguished from the *bona fide* plastid genome. However they are unlikely to represent true plastid DNA as they were isolated from a genomic library resistant to restriction by HpaII, and from the sequence of the tobacco plastid genome these two clones would be expected to contain 22 and 23 HpaII restriction sites respectively.

5.6: **Summary**

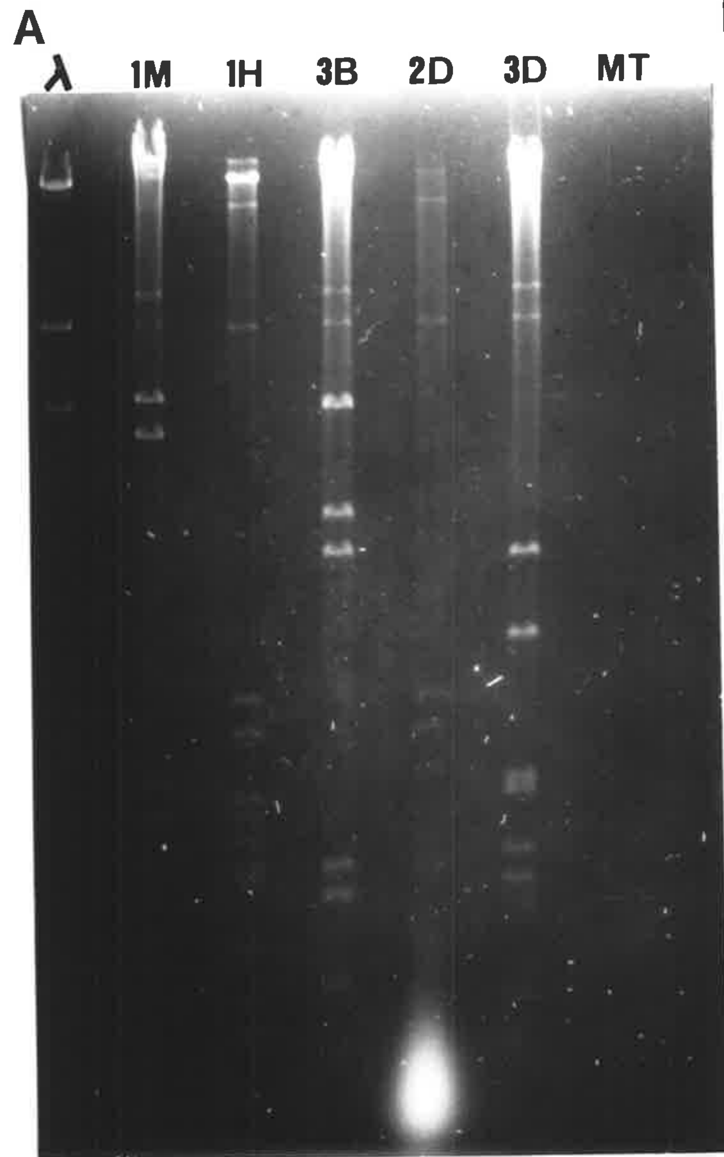
Five lambda clones isolated from a tobacco genomic library resistant to restriction by HpaII were all found, by restriction fragment analysis and sequencing, to be homologous in their entirety to the tobacco plastid genome. Regions of these clones that were sequenced showed greater than 99% sequence homology to the published sequence of the tobacco plastid genome (Shinozaki *et al*, 1986). However sequence

divergence resulted in the inactivation or alteration of essential plastid genes encoded by three of these clones, thereby precluding a true plastid origin of these sequences. From the sequence divergence and the highly methylated state of these cloned sequences it was concluded that they were nuclear in origin. None of these clones were homologous in their entirety to the mitochondrial genome commensurate with a nuclear origin. At least a third of the plastid genome is present within the tobacco nuclear genome as integrants in excess of up to 17.5 kbp. One of these sequences was estimated to be present 15 $\pm$ 5 times within a 1c tobacco nucleus. The large size and multiplicity of these nuclear integrants is consistent with Southern hybridizations shown in chapter 4. The high sequence similarity between these clones and the tobacco plastid genome suggest that they represent evolutionarily recent transfers of plastid DNA to the nuclear genome.

Figure 5.1: Hybridization of tobacco mitochondrial DNA to lambda clones isolated from the tobacco nuclear genomic library.

A) Agarose gel showing EcoRI restriction patterns of DNA isolated from lambda clones 1M, 1H, 3b, 2d and 3D. The track labelled "MT" contains approximately 50 ng of tobacco mitochondrial DNA restricted with EcoRI, while the track labelled '■' contains 1 μg of wild type lambda DNA restricted with HindIII.

B) Autoradiograph showing hybridization of tobacco mitochondrial DNA to a Southern blot of the gel depicted in A).


A
λ
1M
1H
3B
2D
3D
MT

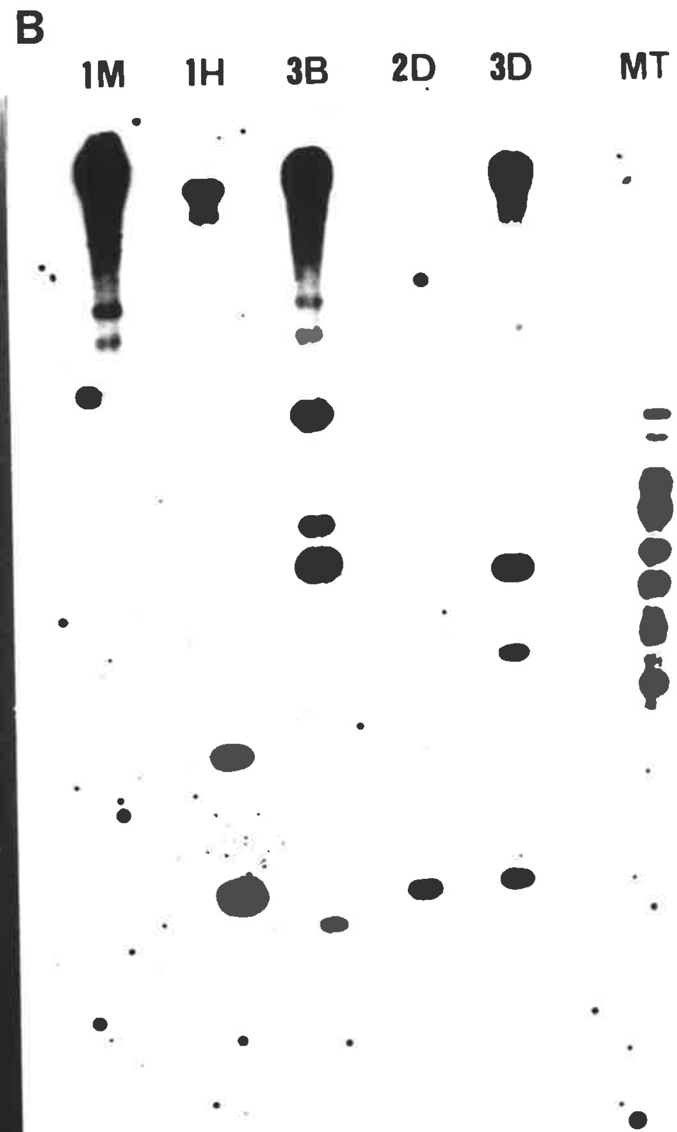
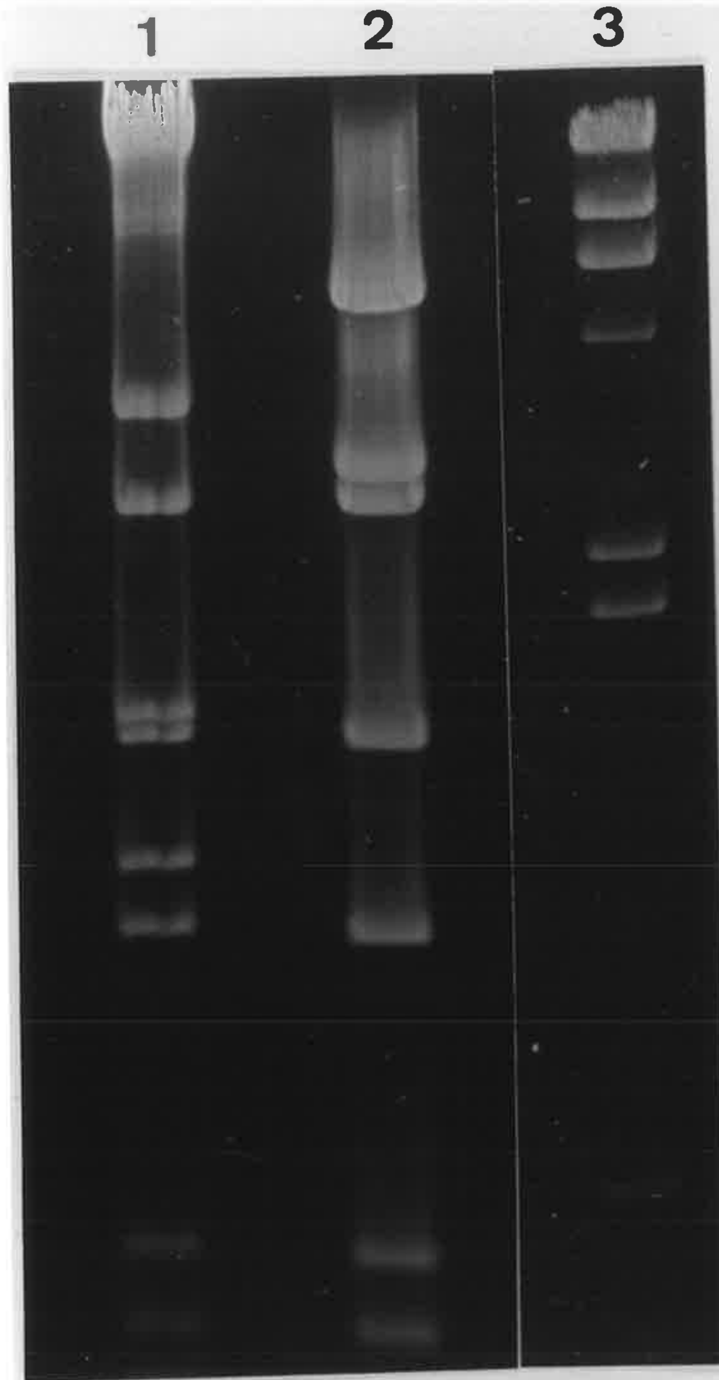

B
1M
1H
3B
2D
3D
MT

Figure 5.2: Hybridization of pTB19 to lambda clone 3D.

A) Agarose gel showing EcoRI restriction patterns of lambda clone 3D DNA (1) and of pTB19 (2) which encompasses the tobacco plastid genome from nucleotide 43,426 to 55,921 (Shinozaki *et al*, 1986). Track 3 contains 1 µg of wild type lambda DNA restricted with HindIII.

B) Hybridization of pTB19 to a Southern blot of the agarose gel depicted in A). The smallest 3 hybridizing bands of clone 3D and pTB19 are a longer exposure of the hybridized filter, compared to the remaining bands. Arrows identify EcoRI restriction fragments of clone 3D that were subcloned and sequenced.


A
1
2
3

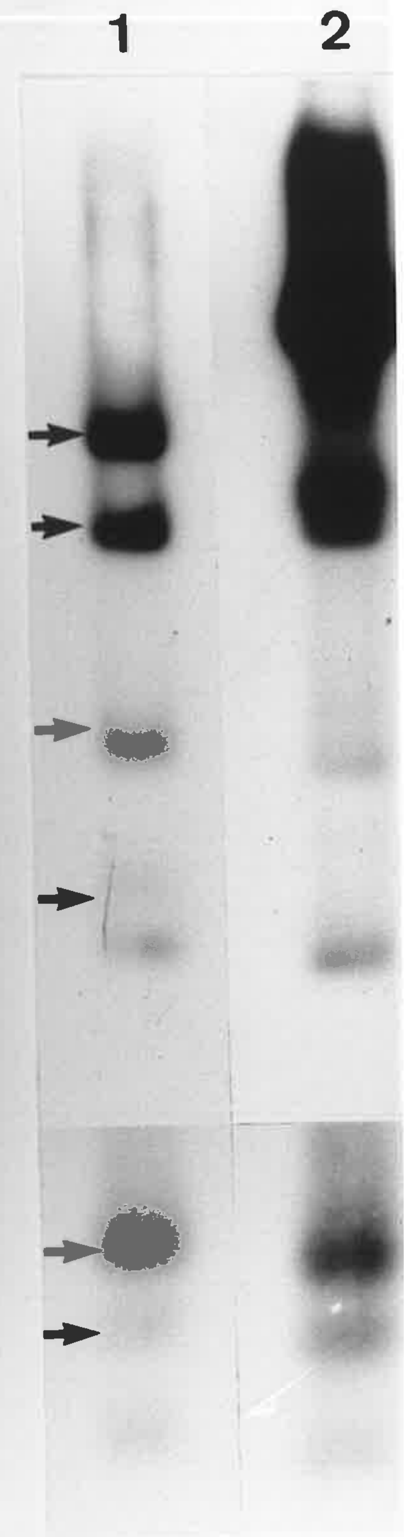

B
1
2

Figure 5.3: EcoRI restriction map of lambda clone 3D.

The tobacco DNA insert of clone 3D is represented by a green line while bacteriophage vector sequences are depicted in red. EcoRI restriction sites are marked with an "E" and the size of restriction fragments resulting from EcoRI restriction of this clone given in kbp. The boxed EcoRI restriction site is unique to clone 3D when compared to the tobacco pt genome from nucleotide 42455-56290 (Shinozaki *et al*, 1986). Plastid genes encoded by clone 3D are boxed in dark green and appropriately labelled. Genes which are surronded by a bracket (eg. *trn*V, *trn*L) contain introns and each exon of the gene is represented as an individual green box. Regions of the clone that have been sequenced are indicated by red arrows. Regions 1-4 are discussed in the text and represent portions of the clone that have been bidirectionally sequenced.

14 kbp

42455 56290

psaA trnS rps4 trnT trnL trnF psbG ndhC trnV trnM atpE/B

E E E E E E E E E E E

3.8 1.6 1.3 1.6 0.3 1.0 0.5 0.2 0.4 2.88

Regions sequenced

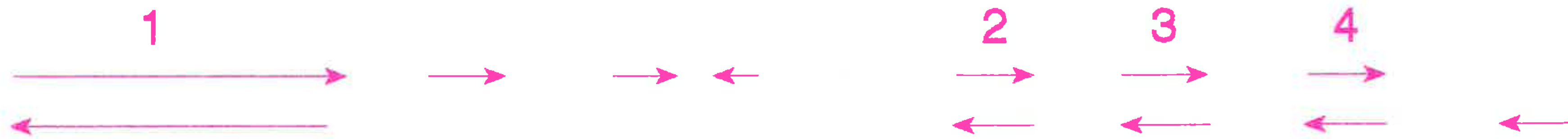


Figure 5.4: DNA sequence homology between region 1 of lambda clone 3D and the tobacco plastid genome.

(Region 1 of clone 3D is illustrated in figure 5.3).

The sequence of the tobacco plastid genome is represented by the top line of sequence labelled "cp". Numbering above the plastid sequence correspond to that of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986). The sequence of region 1 of clone 3D is represented by the bottom line labelled "3D" and numbered such that nucleotide 1-10 correspond to the cloning site of the left arm of the lambda vector (see figure 5.3). Nucleotides of clone 3D that are identical to the published plastid sequence are indicated by a "-", whereas the position of deleted nucleotides in the 3D sequence are marked with "■".

```
        42430     42440     42450     42460     42470     42480
cp    GAGATAATTGAGCATGCCATGACGTTGTTAGGATCTCATATAGGCCTTTATGGCCCTGAC
3D                         GAATTCCCCG----------------------■------
                                   10        20        30

        42490     42500     42510     42520     42530     42540
cp    CTGTAAATGGACCTTTATGAGCTTCTAAAATATCTTTTAGTCCATGACCAATACCCCAGT
3D    ------------------------------------------------------------
       40        50        60        70        80        90

        42550     42560     42570     42580     42590     42600
cp    TGGTCCTATACATGTGACCCGCTATCAGGAAAAGAATTGCAATAGCTAAATGGTGATGGG
3D    ------------------------------------------------------------
      100       110       120       130       140       150

        42610     42620     42630     42640     42650     42660
cp    CAATATCAGTCAGCCACAGACCCCCAGTTACTGGATCTAATCCTCCACGAAAAGTAAGAA
3D    ------------------------------------------------------------
      160       170       180       190       200       210

        42670     42680     42690     42700     42710     42720
cp    AGTCCGCATATTTTGACCAATTCAAGGTGAAAAATGGGGTTGCTCCCTCGGCAAAACTGG
3D    ------------------------------------------------------------
      220       230       240       250       260       270

        42730     42740     42750     42760     42770     42780
cp    GATAAAGTTGAGCCAAAAGATCTCGATTCAAGATAAATTCATGAGGAAGTGGTATCTCTT
3D    ------------------------------------------------------------
      280       290       300       310       320       330

        42790     42800     42810     42820     42830     42840
cp    TAGGATCTACTCCAGCGTTTAGAAATTGGTTAATCGGTAAAGATACATGTACTTGATGCC
3D    ------------------------------------------------------------
      340       350       360       370       380       390

        42850     42860     42870     42880     42890     42900
cp    CCGCCCAAGAGAGAGACCCAAGTCCTAGTAGCCCTGCCAAATGGTGATTCAGCATAGATT
3D    ------------------------------------------------------------
      400       410       420       430       440       450

        42910     42920     42930     42940     42950     42960
cp    CTACATCTTGAAACCAAGCCAATTTTGGCGCCGCTTTATGATAATGAAACCAACCAGCAA
3D    ----------------------------T-------------------------------
      460       470       480       490       500       510

        42970     42980     42990     43000     43010     43020
cp    AAAGCATTAACGCTGCAAAGACCAATGCCCCAATTGCTGTACAATAGAGTTGTAATTCAC
3D    ------------------------------------------------------------
      520       530       540       550       560       570

        43030     43040     43050     43060     43070     43080
cp    TAGTTATTCCAGATGCTCGCCAAATCTGAAAAAAACCAGAGGTTATTTGTATTCCTCGGA
3D    -------------------------T----------------------------------
      580       590       600       610       620       630
```

```
        43090     43100     43110     43120     43130     43140
cp    AACCCCCGCCTACGTCACCATTTAATATTTCTTGGCCCACTATTGGCCAAACCACCTGGG
3D    ------------------------------------------------------------
     640       650       660       670       680       690

        43150     43160     43170     43180     43190     43200
cp    CACTAGGCCCAATGTGAGTTGGATCACTTAGCCACGCTTCATAATTAGAAAAACGAGCAC
3D    -------■--------T-------------------------------------------
     700         710       720       730       740       750

        43210     43220     43230     43240     43250     43260
cp    CGTGGAAATACATGCCGCTCAGCCAAAGAAAGATGATGGAGAGTTGACCGAAATGTGCAC
3D    ------------------------------------------------------------
      760       770       780       790       800       810

        43270     43280     43290     43300     43310     43320
cp    TAAATACTTTTCGAGAGATCTCCTCCAAATCACTGGTATGGCTATCGAAATCGTGAGCAT
3D    ------------------------------------------------------------
      820       830       840       850       860       870

        43330     43340     43350     43360     43370     43380
cp    CAGCATGTAGGTTCCAGATCCAAGTGGTAGTATCAGGCCCTTTAGCTATTGTTCTTGAGA
3D    ------------------------------------------------------------
      880       890       900       910       920       930

        43390     43400     43410     43420     43430     43440
cp    AATGACCCGGTCTGGCCCATTCCTCGAACGAAGTTTTTACGGGATCCCTATCTACCAAAA
3D    ------------------------------------------------------------
      940       950       960       970       980       990

        43450     43460     43470     43480     43490     43500
cp    TTTTAACTTCTGGTTCCGGCGAACGAATAATCATTGAGTCCTCCTCTTTCCGGACAACAC
3D    ------------------------------------------------------------
     1000      1010      1020      1030      1040      1050

        43510     43520     43530     43540     43550     43560
cp    ATACAAAGAGACCCGCCAACAGTCAAATAATTAGTGAACCTTAGAGATAGAGAGATATTT
3D    ---------------------A--------------------------------------
     1060      1070      1080      1090      1100      1110

        43570     43580     43590     43600     43610     43620
cp    CTATAATTAGTTCGTTTCTCTTCTATTTTTCTATCTCCCATCTATCTATTTTCTTTAGTT
3D    ------------T-----------------------------------------------
     1120      1130      1140      1150      1160      1170

        43630     43640     43650     43660     43670     43680
cp    ATTTACTAGAGCAATTATGATCTGGAAGTCGATCCGGGGCAAGTGTTCGGATCTATTATG
3D    ------------------------------------------------------------
     1180      1190      1200      1210      1220      1230

          43690     43700     43710     43720     43730     43740
cp    ACATA-GCCTTGAGGCGCTCAACGGACCTTTTAACCTTCTAAAAACCTTTTTGGGCTTTG
3D    -----T----------------------C----------------------G--------
     1240      1250      1260      1270      1280      1290
```

```
           43750     43760     43770     43780     43790     43800
cp     GATTGATCCAAAAACGACTTTTTTGTGCAACCTAGTGTATATTCATAGAAGTTATTAGAT
3D     ---------------A--------------------------------------------
      1300      1310      1320      1330      1340      1350

           43810     43820     43830     43840     43850     43860
cp     GGAGCTCTTTAATTTTTTACCTAGAAGATTTTAATTACTCTATTCCAAATCACGCGAGTA
3D     --------------------------------------------------T---------
      1360      1370      1380      1390      1400      1410

           43870     43880     43890       43900     43910     43920
cp     GCCATTAGACATTACTAAGAGACATCCCCGC--TATATATATTTAGTGATTCGAGGGTTT
3D     -------------------------------TA---------------------------
      1420      1430      1440      1450      1460      1470

             43930     43940     43950     43960     43970     43980
cp     ATTTTATTAGTTTTAATAATAAGAATTTTGTTTAATTTAATATAATAAACAAAGTCTATT
3D     ------------------------------------------------------------
      1480      1490      1500      1510      1520      1530

             43990     44000     44010     44020     44030     44040
cp     TTGTACTCTATCTGTGTATCCTTTTTTATTCCTAAAAAATAGCAGATGAAATAGAAGGCT
3D     ------------------------------------------------------------
      1540      1550      1560      1570      1580      1590

             44050     44060     44070     44080     44090     44100
cp     TAGAAGGGAGATAATGAAATTATGTGATTGGGTCTTCCAAAAGCAAAGGAATAATCCGTT
3D     ----------------------------------------------------G-------
      1600      1610      1620      1630      1640      1650

             44110     44120     44130     44140     44150     44160
cp     TTTTAGTTAACTGATCTGATGGGTCCAACAAACAATAAATTATAACAAATATCTAAATTC
3D     ------------------------------------------------------------
      1660      1670      1680      1690      1700      1710

             44170     44180     44190     44200     44210     44220
cp     TAAATAAAAAAATCAAAAATAATAGACTAAGATTCTAAATAAAGGATAATAAATAAACGG
3D     --------------C---------------------------------------------
      1720      1730      1740      1750      1760      1770

             44230     44240     44250     44260     44270     44280
cp     GATCTTCTTTTATTCGAAACGTCTCGTGATCTTCAACCAATTATGCGCTTCAATATAATT
3d     ------------------------------------------------------------
      1780      1790      1800      1810      1820      1830

             44290     44300     44310     44320     44330     44340
cp     ACCGGGAGTAAGCGCTATAGCCTGTTTCCAATACTCAGCGGCTTGATCGAACCAAGCCTC
3D     ------------------------------------------------------------
      1840      1850      1860      1870      1880      1890

             44350     44360     44370     44380     44390     44400
cp     TGCAATTTCAGAATCTCCCTGTTGAATGGCCTGTTCTCCCCGGCCGGAATAGGTAGTTCA
3D     ------------------------------------------------------------
      1900      1910      1920      1930      1940      1950
```

```
         44410     44420     44430     44440     44450     44460
cp   ATTCCTTCCCTTAGAACCGTACTTGAGAATTTCTTACCTCATACGGCTCAGCAGTCAATT
3D   ------------------------------------------------------------
    1960      1970      1980      1990      2000      2010

         44470     44480     44490     44500     44510     44520
cp   CTTTTGGTGTCCCATTTTGATCTATACCATATCTAATAAAATCTAATGAGATTTCTCATG
3D   ------------------------------------------------------------
    2020      2030      2040      2050      2060      2070

         44530     44540     44550     44560     44570     44580
cp   GATCTATCCCAGTTTTAGGGTTAACCAAAAGCCAAATAGGTTAATTACATGAGTTTCAAA
3d   -----------------A-------------AA---------------------------
    2080      2090      2100      2110      2120      2130

         44590     44600     44610     44620     44630     44640
cp   CTGAAATTTGGATGAATAATCCGTTTATTTAGTTTTATCTTTTTTCCCACCTTCAGAAGA
3D   ------------------------------------------------------------
    2140      2150      2160      2170      2180      2190

         44650     44660     44670     44680     44690     44700
cp   ATAAAGCATAGGCATTTCTACTAGTGTTAGAATTTTATGAAAGGTAACTATCTCGGTTTC
3D   ------------------------------------------------------------
    2200      2210      2220      2230      2240      2250

         44710     44720     44730     44740     44750     44760
cp   ATAGATAAATTTATATAGAATCTTTGAAAAAGACTTTCTTTCATAAGAAAGAAAATACTT
3D   ------------------------------------------------------------
    2260      2270      2280      2290      2300      2310

         44770     44780     44790     44800     44810     44820
cp   ACTATCTTTGGGATCTGATCCTACACCGCTGCTCAAGACTTTAGTGGATCGACTCTATTA
3D   ------------------------------------------------------------
    2320      2330      2340      2350      2360      2370

         44830     44840     44850     44860     44870     44880
cp   CATAAGTTAATTCCTAATTTTTATTTCACATCATGAGATAAGTATTTCTTCCATCATGAC
3D   ------------------------C-----------------------------------
    2380      2390      2400      2410      2420      2430

         44890     44900     44910     44920     44930     44940
cp   ATAAGTACGCAGTTATTATTGTATCGGCCCAAAACCTCGCTAATTGATCTTTACGGTGCT
3D   ------------------------------------------------------------
    2440      2450      2460      2470      2480      2490

         44950     44960      44970     44980     44990     45000
cp   TCCTCTATCTCTATCAATTAAAGCCTT-ATATCCATAGAAAAAAGTTGCTAGGCATTTTT
3D   ---------------------------A--------------------------------
    2500      2510      2520      2530      2540      2550

          45010     45020     45030     45040     45050     45060
cp   ATTTTTTCCTATTTTGACTTCTATGAAGTTTCTTTCTTTGCTACAGCTGATAAAAATCGT
3D   ------------------------------------------------------------
    2560      2570      2580      2590      2600      2610
```

```
         45070     45080     45090     45100     45110     45120
cp  TGTTTTAGACGATGCATATGTAGAAAGCCTATTTGGTTCTACTAGTTACTTTACTAGATT
3D  -----------------------------------------G------------------
    2620      2630      2640      2650      2660      2670

         45130     45140     45150     45160     45170     45180
cp  TTTCTTTTTTTTTTTTTTTTTCTTTCTATAGTGGAGATAGTCGCACGTAATGACAGATCA
3D  ----██████--------------------------------------------------
    2680           2690      2700      2710      2720      2730

         45190     45200     45210     45220     45230     45240
cp  CGGCCATATTATTAAAAGCTTGTGGTAAGAATGGGTTTCGTTCTAGTGCTCGAAAATAAT
3D  ------------------------------------------------------------
         2740      2750      2760      2770      2780      2790

         45250     45260     45270     45280     45290     45300
cp  ATTCCAAAGCTTTCGTATGTTCTCCATTACTTGTGTGGATAAGCCCTATATTATAGAGTA
3D  ------------------------------------------------------------
         2800      2810      2820      2830      2840      2850

         45310     45320     45330     45340     45350     45360
cp  TATAACTTCGATCATAGGGATCAATTTCTAGTCGCATAGCTTCATAATAATTCTGCAAAG
3D  ---------------------------------------------------------G--
         2860      2870      2880      2890      2900      2910

         45370     45380     45390     45400     45410     45420
cp  CTTCCGCGTAATTTCCTTCGGATTGAGCCGACATCCGTTACGGTCGTCATTCAATTGAAA
3D  ------------------------------------------------------------
         2920      2930      2940      2950      2960      2970

         45430     45440     45450     45460     45470     45480
cp  GAATCTCCGTTCCAGAACCGTACGTGAGATTTTCACCTCATACGGCTCCTCCCTTATGTG
3D  -----------------
         2980
```

Figure 5.5: Amino acid sequence of the plastid *psa*A gene encoded by lambda clone 3D.

The tobacco plastid DNA sequence encoding the amino terminus of the *psa*A protein is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Immediately beneath this plastid sequence is the homologous region of clone 3D which is numbered as in figure 5.4. Nucleotides common to both the cp and 3D sequence are marked with "-" in the 3D sequence. The positions of deleted nucleotides in the 3D sequence are marked with '■'. The amino acid chain produced by translation of the 3D sequence occurs directly beneath it.

```
        43470
cp  ATGATTATTCGTTCGCCGGAACCAGAAGTTAAAATTTTGGTAGATAGGGATCCCGTAAAA
3D  ---------+---------+---------+---------+---------+---------+
    MetIleIleArgSerProGluProGluValLysIleLeuValAspArgAspProValLys -
            1020

cp  ACTTCGTTCGAGGAATGGGCCAGACCGGGTCATTTCTCAAGAACAATAGCTAAAGGGCCT
3D  ---------+---------+---------+---------+---------+---------+
    ThrSerPheGluGluTrpAlaArgProGlyHisPheSerArgThrIleAlaLysGlyPro -

cp  GATACTACCACTTGGATCTGGAACCTACATGCTGATGCTCACGATTTCGATAGCCATACC
3D  ---------+---------+---------+---------+---------+---------+
    AspThrThrThrTrpIleTrpAsnLeuHisAlaAspAlaHisAspPheAspSerHisThr -

cp  AGTGATTTGGAGGAGATCTCTCGAAAAGTATTTAGTGCACATTTCGGTCAACTCTCCATC
3D  ---------+---------+---------+---------+---------+---------+
    SerAspLeuGluGluIleSerArgLysValPheSerAlaHisPheGlyGlnLeuSerIle -

cp  ATCTTTCTTTGGCTGAGCGGCATGTATTTCCACGGTGCTCGTTTTTCTAATTATGAAGCG
3D  ---------+---------+---------+---------+---------+---------+
    IlePheLeuTrpLeuSerGlyMetTyrPheHisGlyAlaArgPheSerAsnTyrGluAla -

                                                       43120
cp  TGGCTAAGTGATCCAACTCACATTGGGCCTAGTGCCCAGGTGGTTTGGCCAATAGTGGGCC
3D  ---------+-------A-+------■--+---------+---------+---------+
    TrpLeuSerAspProThrHisIleGly-LeuValProArgTrpPheGlyGlnEnd
                                                  680
```

Figure 5.6: DNA sequence homology between region 2 of lambda clone 3D and the tobacco plastid genome.

(Region 2 of clone 3D is illustrated in figure 5.3)

The DNA sequence of the tobacco plastid genome is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Nucleotides common to both the cp sequence and region 2 of clone 3D (labelled 3D) are indicated with a "-" in the 3D sequence.

```
      50850     50860     50870     50880     50890     50900
cp  GAATTCCAATGAATCCCTAATTGTCTTTTTTTGTTAGCCTATCGATAATTCCCTAAATTA
3D  ---------C----------------------------------------T---------

      50910     50920     50930     50940     50950     50960
cp  GACCTGCTTAATCTAGAACAGAACGTGCAATCCTTGAATATCTGAAATTGTCTAAGTGGA
3D  ---------------T---------------------------------A----------

      50970     50980     50990     51000     51010     51020
cp  AATAGCTTTCTTATCATTCAATGAGCATCTTGTATTTCATAAAAATTGGGGGCAATATAA
3D  ------------------------------------------------------------

      51030     51040     51050     51060     51070     51080
cp  TCCTTACGTAAGGGCCATCCTATCCAACTTTCAGGCATTAAGATACGTTTCAAGCGTGGA
3D  ------------------------------------------------------------

      51090     51100     51110     51120     51130     51140
cp  TGATTATCATAAGAGATTCCCAACATATCATATGATTCTCGTTCTTGAAAATCCACACTT
3D  ------------------------------------------------------------

      51150     51160     51170
cp  TTCCAAACCCAGAAAACAGACGGAATTC
3D  ----------------------------
```

Figure 5.7: DNA sequence homology between region 3 of lambda clone 3D and the tobacco plastid genome.

(Region 3 of clone 3D is illustrated in figure 5.3)

The DNA sequence of the tobacco plastid genome is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Nucleotides common to both the plastid genome and region 3 of clone 3D (labelled 3D) are indicated by "-" in the 3D sequence.

```
         52210     52220     52230     52240     52250     52260
cp  GTTGTTCGATCAAGTAAAGGAAACTGAATGGAATTCATAACTGTCTCAATCTTATTTTTT
3D                                ------------------------------

         52270     52280     52290     52300     52310     52320
cp  CCGTTTTTCTTTTTATTGTCTGAATATTCAGGAGCTAAGACCATTCCAATGCCCCCTTTC
3D  ------------------------------------------------------------

         52330     52340     52350     52360     52370     52380
cp  GCCATGCATAAACTAAACCAATAATTAAGATAAGCACGAAAATGAAAGCTTCTATAAATA
3D  ------------------------------------------------------------

         52390     52400     52410     52420     52430     52440
cp  CAGATACACCCAATACGTCGAAACTCATTGCCCATGGATAAAGAAAAACCGTTTCAACAT
3D  ------------------------------------------------------------

         52450     52460     52470     52480     52490     52500
cp  CAAAAACAACAAAAACTAGAGCAAACATATAATAACGGATTCGAAATTGTAACCAAGCAT
3D  ------------------------------------------------------------

         52510     52520     52530     52540     52550     52560
cp  CGCCCATTGGTTCTATACCCGACTCATAAGTAGAAAGTTTCTCCGGCCCTTTGCTAATCG
3D  ------------------------------------------------------------

         52570     52580     52590     52600     52610     52620
cp  GGGCTAACACTCCGGAAATTAAAAATGCCAAAATAGGAACAAGGATAGATATTATTAGAA
3D  ----------------------------------------------G-------------

         52630     52640     52650     52660     52670     52680
cp  ATGCCCAAAAAAAATCATATTCGTAAAGCAGAAACATAAACGCACTCCTATGAACGTGGA
3D  ------------------------------------------------------------

         52690     52700     52710     52720     52730     52740
cp  AAATATACCGGATTCGATTGGTCGATTCGAATTGGAATTGTCAAGTCATCCATAACTATT
3D  ---------------------------T--------------------------------

         52750     52760     52770     52780     52790     52800
cp  TAGTCAAAACAAGAATTCATTTTGATCGAACCGTCTAGTTTGCTTTGTTTATTGGTTTAT
3D  ------------------
```

Figure 5.8: Nucleotide sequence comparisions of homologous chloroplast and nuclear DNA

DNA sequence comparisions between the tobacco plastid genome (cp) (Shinozaki *et al*, 1986), the PCR product obtained from total tobacco leaf DNA (pcrt), the PCR product obtained from tobacco nuclear DNA restricted with EcoRV (pcrn) and region 4 of lambda clone 3D (3D). Nucleotides identical to the published tobacco plastid genome sequence (Shinozaki *et al*, 1986) are indicated with a dashed line, while the positions of nucleotide deletions are marked with "■". Direct repeats associated with the 12 and 41 bp deletion in the 3D sequence are shown in bold and underlined with arrows. The position and sequence of PCR primers are labelled "pri".

```
      53400      53410     53420     53430     53440     53450     53460
  cp  CATGTGAATTCCTCTTCGTAGTTTTTCATTTCACCAGGCCCGTGAAATGATTTGACTTCCACAAC
pcrt  -----------------------------------------------------------------
pcrn  ------------------------------------------------------------A--T
  3D       -------------------------------------------------------A--T
 pri  CATGTGAATTCCTCTTCGTA
           Eco RI

            53470     53480     53490     53500     53510     53520
  cp  TCAATAAGATTGGGGATATCAAAAGAAAGGGAGTCTCACTAATTCTTTTATTGTGGATAT
pcrt  ------------------------------------------------------------
pcrn  -------------------██████████████████████████████████████████
  3D  -------------------██████████████████████████████████████████
                    Eco RV

            53530     53540     53550     53560     53570     53580
  cp  GAATATGTAATTCGCCTCCGAAGATTAATGACGAAAGGTTGGTTTCTTTATCCGCAATTG
pcrt  ------------------------------------------------------------
pcrn  ---------------------------------------------G--------------
  3D  ---------------------------------------------G--------------
            53590     53600     53610     53620     53630     53640
  cp  AAAAAATCAATATCGATTGGATCCGTTGATATGCATTTTTTCTTTCATCTGCTTAAACGA
pcrt  ------------------------------------------------------------
pcrn  ----------------------------------G-------------------------
  3D  ----------------------------------G-------------------------
            53650     53660     53670     53680     53690     53700
  cp  TTGCCGTGAGTAAACTTATAGGAATAATTGGATTTCACTTAGTTACAAGCAAGAAATAAT
pcrt  ------------------------------------------------------------
pcrn  ------------------------------------------------------------
  3D  ------------------------------------------------------------
            53710     53720     53730     53740     53750     53760
  cp  AATGAAGAAATGAAAATTATAGAATTTTTTGGATTTTGCATTTTTATAGGGCTATACGGA
pcrt  ------------------------------------------------------------
pcrn  ---------------------C------████████████--------------------
  3D  ---------------------C------████████████--------------------

            53770     53780     53790     53800     53810     53820
  cp  CTCGAACCGTAGACCTTCTCGGTAAAACAGGTCAAACTTATTATTATTAAAATGATCTGA
pcrt  ------------------------------------------------------------
pcrn  ----------------------------G-A---------------------C-------
  3D  ----------------------------G-A------end
            53830     53840
  cp  ACTGTTTCAAAGACCCAACATGC
pcrt  -----------------GA--T-
pcrn  -----------------GA--T-
 pri    ACAAAGTTTCTGGGTCTTAAGTAA 5
                       Eco RI
```

Figure 5.9: Detection and quantification of nuclear genomic sequences with homology to cpDNA using the polymerase chain reaction.

A) Agarose gel showing the PCR product obtained from 1 μg of tobacco nuclear DNA restricted with EcoRV (lane 2). Lane 1 shows that no PCR product was obtained from a control reaction containing no DNA. Lane 3 contains 0.5 μg of wild type lambda DNA restricted with HindIII.

B) A 2% agarose gel showing PCR products used to quantify the number of tobacco nuclear sequences containing 12 and 41 bp deletions compared with plastid DNA. Lanes 2-7 are PCR products obtained from 1 μg of nuclear DNA restricted with EcoRV and having 0, 2, 5, 10, 20 and 100 copies respectively of a plasmid construct added per 1c tobacco nuclear genome. Lane 8 is the product obtained from a PCR reaction containing only construct plasmid and lane 9 is a control showing no PCR products from a reaction with neither plasmid nor nuclear DNA added. Lanes 1 and 10 are PCR products that have been cloned into pUC19 and sequenced, with the 397 bp fragment in lane 1 being identical in sequence to region 4 of lambda clone 3D and the 450 bp fragment in lane 10 being identical to the published tobacco plastid genome from nucleotide 53,396 to 53,846 (Shinozaki *et al*, 1986). Lane 11 contains 1 μg HindIII restricted lambda DNA.

(a) 1 2 3

(b) 1 2 3 4 5 6 7 8 9 10 11

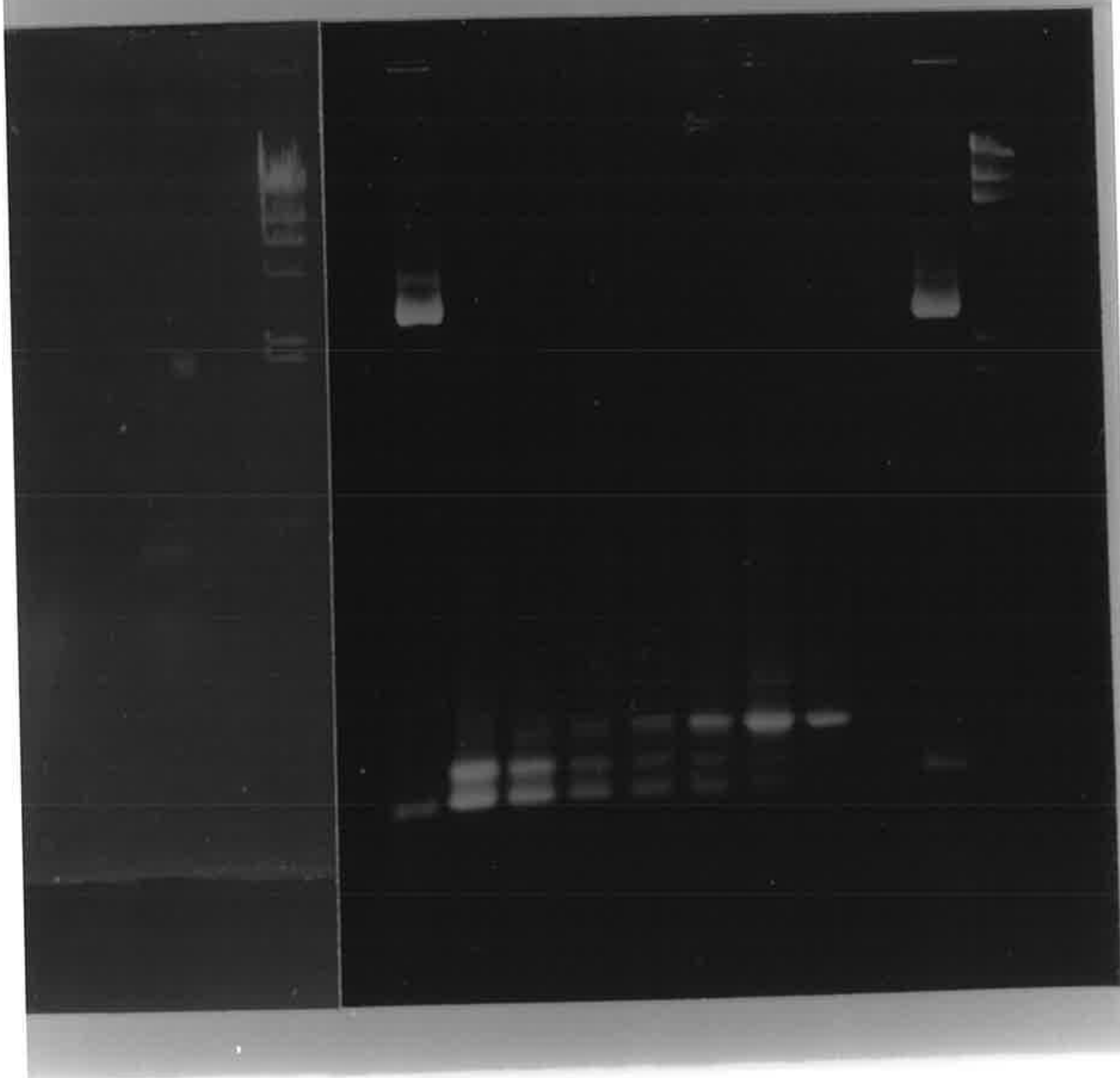


Figure 5.10: PCR analysis of the tobacco mitochondrial genome for homology to the tobacco plastid genome from nucleotide 53,396 to 53,846 (Shinozaki *et al*, 1986).

2% agarose gel showing the PCR products obtained from;

1) 1 μg of total tobacco DNA restricted with EcoRV.
2) 1 μg of total tobacco DNA restricted with EcoRV and with 24 copies of plasmid construct added per 1c tobacco nuclear genome.
3) 100 ng of tobacco mitochondrial DNA restricted with EcoRV and with 1 plasmid construct added per mt genome.
4) 100 ng of unrestricted mitochondrial DNA with 1 copy of plasmid construct added per mitochondrial genome.
5) Plasmid construct alone; equivalent to the amount of plasmid DNA added to the PCR reactions whose products are shown in lanes 3 and 4.
6) 100 ng of tobacco mitochondrial DNA.

Lane 7 contains 1 μg of wild type lambda DNA restricted with HindIII. No PCR products were obtained from a PCR reaction with no DNA added (not shown). PCR primers were identical to those depicted in figure 5.8.

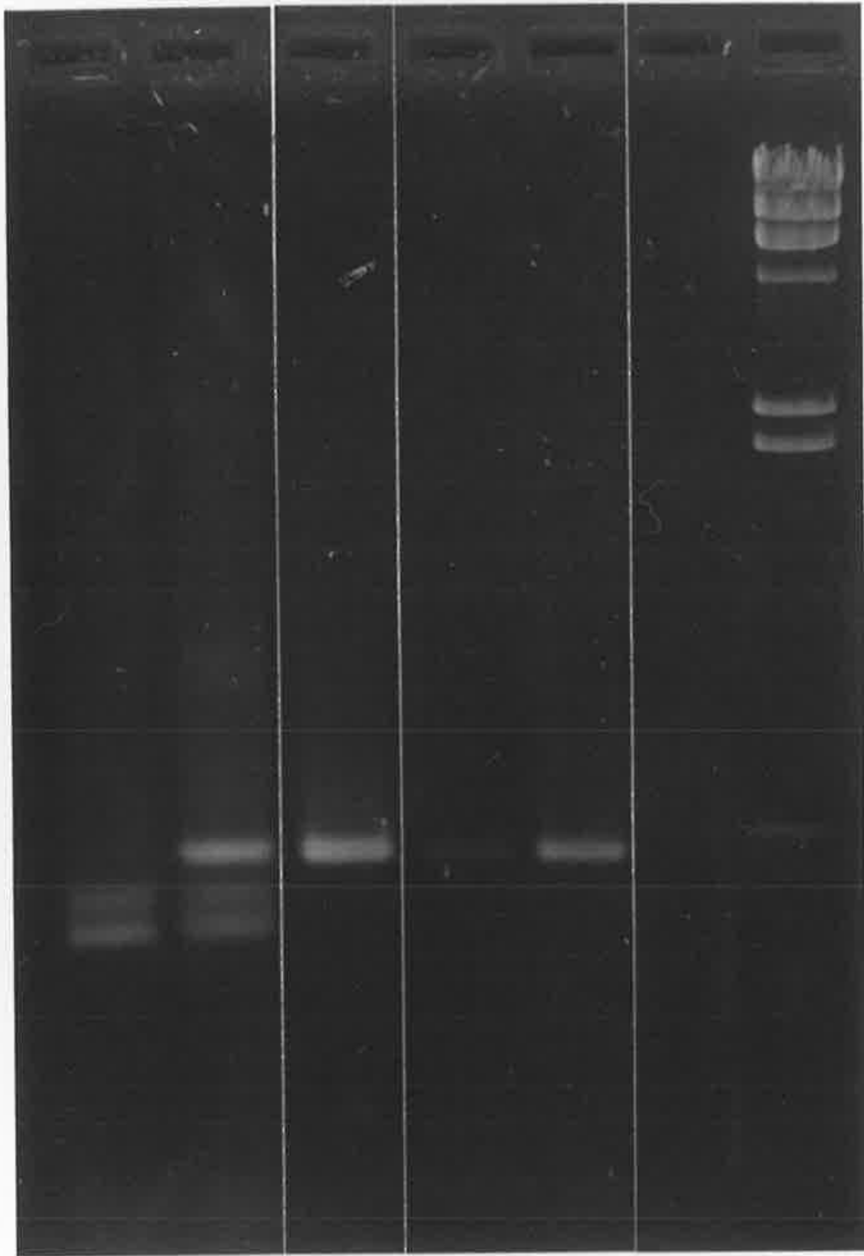

1 2 3 4 5 6 7

Figure 5.11: Restriction enzyme mapping of lambda clone 3b.

A) Expected size restriction fragments obtained from a lambda EMBL4 clone containing a DNA insert encompassing nucleotide 24,962-42791 of the tobacco plastid genome (Shinozaki *et al*, 1986), upon restriction with EcoRI, EcoRI + XbaI and XbaI. L and R refer to the left and right arms of the lambda vector which are 19.9 and 8.8 kbp in size respectively.

B) Agarose gel showing restriction products of lambda clone 3b DNA upon digestion with 1) EcoRI, 2) EcoRI + XbaI and 3) XbaI. Track 4 contains 0.5 μg of wild type lambda DNA restricted with HindIII.

A

EcoRI	EcoRI-XbaI	XbaI
L	L	L+8.9
R	R	R+0.8
6.6	5.0	6.9
4.5	4.5	0.77
3.9	3.9	0.37
1.2	1.55	
1.0	0.8	
0.6	0.77	
	0.6	
	0.26	
	0.21	
	0.16	

B

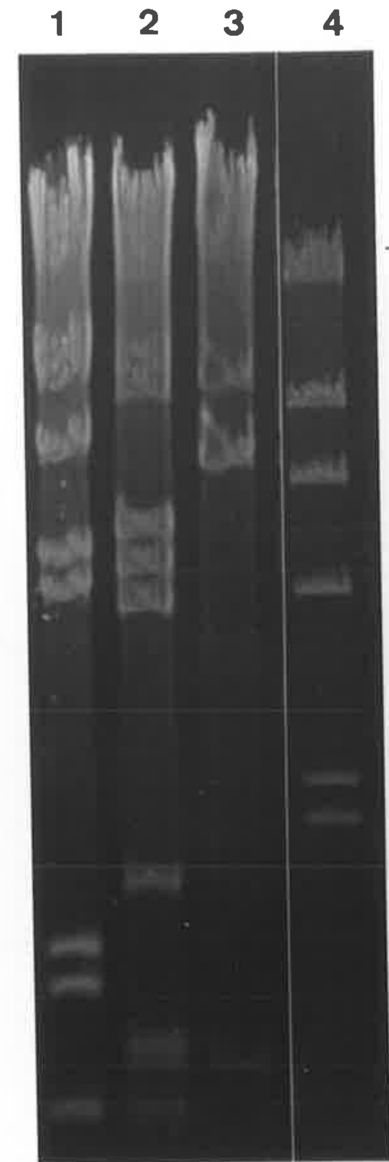
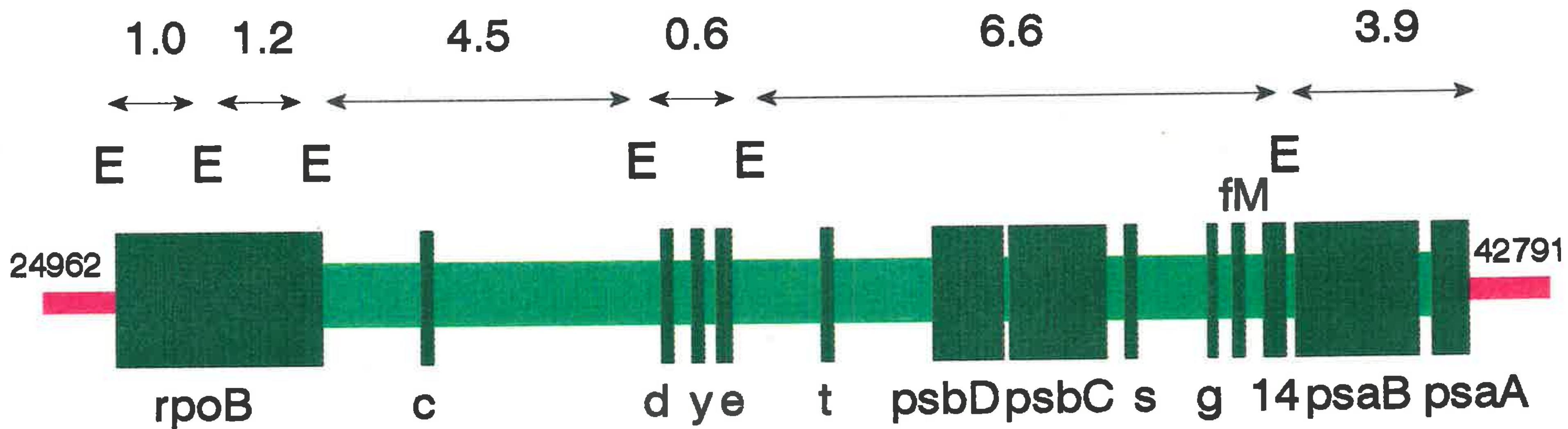
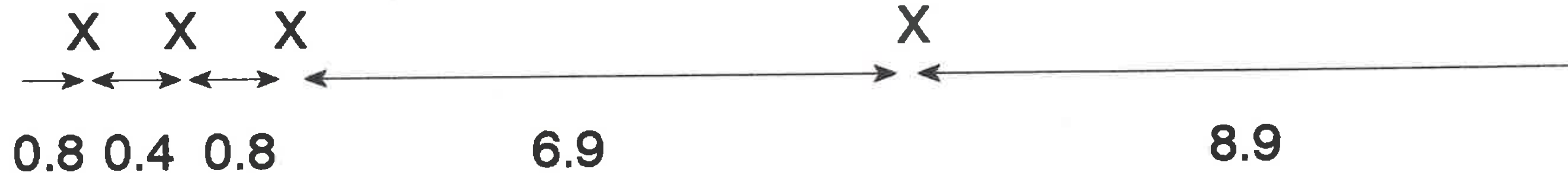


Figure 5.12: Restriction enzyme map of lambda clone 3b.

The tobacco DNA insert of lambda clone 3b is represented by a light green line while vector arms are depicted as a thin red line. EcoRI and XbaI restriction sites are marked with an"E" and "X" respectively. Restriction fragments produced by digestion of clone 3b with these two enzymes are indicated in kbp. Clone 3b is homologous to the tobacco plastid genome from nucleotide 24,962-42,791 (Shinozaki *et al,* 1986) and plastid genes encoded by this clone are boxed in dark green and appropriately labelled. Genes labelled with a single lower case letter are tRNA genes. Regions of clone 3b that have been sequenced are marked with purple arrows. Region 1, 2 and 3 are discussed in the text. Region 3 (which is unlabelled) encompasses the 1.0 kbp EcoRI restriction fragment.



1.0
1.2
4.5
0.6
6.6
3.9
E
E
E
E
E
E
fM
24962
42791
rpoB
c
d
y
e
t
psbD
psbC
s
g
14
psaB
psaA
regions sequenced
2
1



X
X
X
X
0.8
0.4
0.8
6.9
8.9

Figure 5.13: DNA sequence homology between the tobacco plastid genome and region 1 of lambda clone 3b.

(Region 1 of clone 3b is illustrated in figure 5.12)

The sequence of the tobacco plastid genome is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Nucleotides common to the plastid genome and region 1 of clone 3b (labelled "3b") are indicated with an "-" in the 3b sequence, while the positions of nucleotides deleted in the 3b sequence are marked with '■'. The last 10 nucleotides of the 3b sequence correspond to the lambda EMBL4 cloning site and therefore show no homology to the plastid genome.

```
    42410     42420     42430     42440     42450     42460
cp  CCTAACATAGCTAAGTTAAGAGATAATTGAGCATGCCATGACGTTGTTAGGATCTCATAT
3b                                ------------------------------

    42470     42480     42490     42500     42510     42520
cp  AGGCCTTTATGGCCCTGACCTGTAAATGGACCTTTATGAGCTTCTAAAATATCTTTTAGT
3b  ------------------------------------------------------------

    42530     42540     42550     42560     42570     42580
cp  CCATGACCAATACCCCAGTTGGTCCTATACATGTGACCCGCTATCAGGAAAAGAATTGCA
3b  ------------------------------------------------------------

    42590     42600     42610     42620     42630     42640
cp  ATAGCTAAATGGTGATGGGCAATATCAGTCAGCCACAGACCCCCAGTTACTGGATCTAAT
3b  ------------------------------------------------------------

    42650     42660     42670     42680     42690     42700
cp  CCTCCACGAAAAGTAAGAAAGTCCGCATATTTTGACCAATTCAAGGTGAAAAATGGGGTT
3b  -----------------------------------■------------------------

    42710     42720     42730     42740     42750     42760
cp  GCTCCCTCGGCAAAACTGGGATAAAGTTGAGCCAAAAGATCTCGATTCAAGATAAATTCA
3b  ------------------------------------------------------------

    42770     42780     42790     42800     42810     42820
cp  TGAGGAAGTGGTATCTCTTTAGGATCTACTCCAGCGTTTAGAAATTGGTTAATCGGTAAA
3b  --------------------------CGGGGAATTC
```

Figure 5.14: Open reading frames of the plastid *psa*A gene encoded by clone 3b.

The sequence of a region of the tobacco plastid *psa*A gene is represented by the upper line of the sequence (cp) and numbered according to the published sequence of the tobacco plastid genome (Shinozaki *et al*, 1986). The homologous region of clone 3b is positioned immediately beneath the cp sequence and labelled 3b. Nucleotides common to the cp and 3b sequence are indicated by "-" while deleted nucleotides are represented by "■" in the 3b sequence. The amino acid sequence encoded by the three potential open reading frames of this region of clone 3b are labelled a, b and c. The position of translation termination codons in these amino acid sequences are marked with an asterix.

```
         42790
cp       GATCCTAAAGAGATACCACTTCCTCATGAATTTATCTTGAATCGAGATCTTTTGGCTCAA
3b       ---------+---------+---------+---------+---------+---------+

a        D  P  K  E  I  P  L  P  H  E  F  I  L  N  R  D  L  L  A  Q  -
b         I  L  K  R  Y  H  F  L  M  N  L  S  *  I  E  I  F  W  L  N -
c          S  *  R  D  T  T  S  S  *  I  Y  L  E  S  R  S  F  G  S  T-

cp       CTTTATCCCAGTTTTGCCGAGGGAGCAACCCCATTTTTCACCTTGAATTGGTCAAAATATG
3b       ---------+---------+---------+---------+---------+■---------+

a        L  Y  P  S  F  A  E  G  A  T  P  F  F  T  L  N  C  Q  N  M  -
b         F  I  P  V  L  P  R  E  Q  P  H  F  S  P  *  I  V  K  I  C -
c          L  S  Q  F  C  R  G  S  N  P  I  F  H  L  E  L  S  K  Y  A-

cp       CGGACTTTCTTACTTTTCGTGGAGGATTAGATCCAGTAACTGGGGGTCTGTGGCTGACTG
3b       ---------+---------+---------+---------+---------+---------+

a        R  T  F  L  L  F  V  E  D  *  I  Q  *  L  G  V  C  G  *  L
b         G  L  S  Y  F  S  W  R  I  R  S  S  N  W  G  S  V  A  D  *
c          D  F  L  T  F  R  G  G  L  D  P  V  T  G  G  L  W  L  T  D-

cp       ATATTGCCCATCACCATTTAGCTATTGCAATTCTTTTCCTGATAGCGGGTCACATGTATA
3b       ---------+---------+---------+---------+---------+---------+

a        I  L  P  I  T  I  *  L  L  Q  F  F  S  *  *  R  V  T  C  I  -
b         Y  C  P  S  P  F  S  Y  C  N  S  F  P  D  S  G  S  H  V  * -
c          I  A  H  H  H  L  A  I  A  I  L  F  L  I  A  G  H  M  Y  R-

cp       GGACCAACTGGGGTATTGGTCATGGACTAAAAGATATTTTAGAAGCTCATAAAGGTCCAT
3b       ---------+---------+---------+---------+---------+---------+

a        G  P  T  G  V  L  V  M  D  *  K  I  F  *  K  L  I  K  V  H  -
b         D  Q  L  G  Y  W  S  W  T  K  R  Y  F  R  S  S  *  R  S  I -
c          T  N  W  G  I  G  H  G  L  K  D  I  L  E  A  H  K  G  P  F-

                                                     42440
cp       TTACAGGTCAGGGCCATAAAGGCCTATATGAGATCCTAACAACGTCATGGCATGC
3b       ---------+---------+---------+---------+---------+-----

a        L  Q  V  R  A  I  K  A  Y  M  R  S  *  Q  R  H  G  M     -
b         Y  R  S  G  P  *  R  P  I  *  D  P  N  N  V  M  A  C    -
c          T  G  Q  G  H  K  G  L  Y  E  I  L  T  T  S  W  H      -
```

Figure 5.15: DNA sequence comparision between the tobacco plastid genome and region 2 of lambda clone 3b.

(Region 2 of clone 3b is illustrated in figure 5.12)

The DNA sequence of the tobacco plastid genome is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Nucleotides common to the plastid genome and region 2 of clone 3b (labelled "3b") are indicated with "-", while nucleotides deleted in the 3b sequence are represented by '■'.

```
         31640     31650     31660     31670     31680
cp  TTATATTAAATTTTGTATTGGACAAGAAAGGAATTCCCCTTGTGTATGCGCGCCTCAAAA
3b                                ------------------------------

         31700     31710     31720     31730     31740
cp  AGGTATAGTACTCGATTCCATTACATGCATCGGGGGCAATCGAAAAAGCCAGCATTTCTT
3b  ------------------------------------------------------------

         31760     31770     31780     31790     31800
cp  GGAATACTGACTATAATGCTACCAATAATCGTACTAATCCAACCGCATATGTCTTTCTCC
3b  ------------------------------------------------------------

         31820     31830     31840     31850     31860
cp  TACCAAAAGGAAAGAAAAAAGAAATAAGGATTTCCCCTTTGCTTTGACAATGAAATTCTG
3b  ------------------------------------------------------------

         31880     31890     31900     31910     31920
cp  CCCCCGGTCCCCTTCATAAAAAGGGAGAGATTTATTGATATATTTATTGGATCCATCGGG
3b  ------------------------------------------------------------

         31940     31950     31960     31970     31980
cp  ACTGACGGGGCTCGAACCCGCAGCTTCCGCCTTGACAGGGCGGTGCTCTGACCAATTGAA
3b  ------------------------------------------------------------

         32000     32010     32020     32030     32040
cp  CTACAATCCCAGGGAAATACGGGATCTAGCAGAAAATTTGATTCTTTTTTATCTCCGGAT
3b  ------------------------------------------------------------

         32060     32070     32080     32090     32100
cp  CGGGTATTTCTGAAGTACGAAGGGGGTTATATCATCTCATGGCGGATTGGCGAATTTTTG
3b  ------------------A-----------------------------------------

         32120     32130     32140     32150     32160
cp  GGCCGAGCTGGATTTGAACCAGCGTAGACATATTGCCAACGAATTTACAGTCCGTCCCCA
3b  ------------------------------------------------------------

         32180     32190     32200     32210     32221
cp  TTAACCGCTCGGGCATCGACCCAAGAAGAATCAATTTTAGACTTATTGGTAATCCATGAT
3b  -----------------------G------------------------------------

         32240     32250     32260     32270     32280
cp  CAACTTCCTTTCGTAGTACCCTACCCCCAGGGGAATTCGAATCCCCGCTGCCTCCTTGAA
3b  -----------------------------■--------
```

Figure 5.16: Restriction enzyme mapping of lambda clone 3-2

A) Expected size restriction fragments from a lambda EMBL4 clone containing a DNA insert encompassing nucleotide 28,760 to approximately 46,300 of the tobacco plastid genome (Shinozaki *et al*, 1986), upon restriction with BamHI, BamHI + EcoRI, EcoRI, EcoRI + XbaI, XbaI, SacI + XbaI and SacI. L and R refer to the left and right arms of the lambda vector which are 19.9 and 8.8 kbp in size respectively.

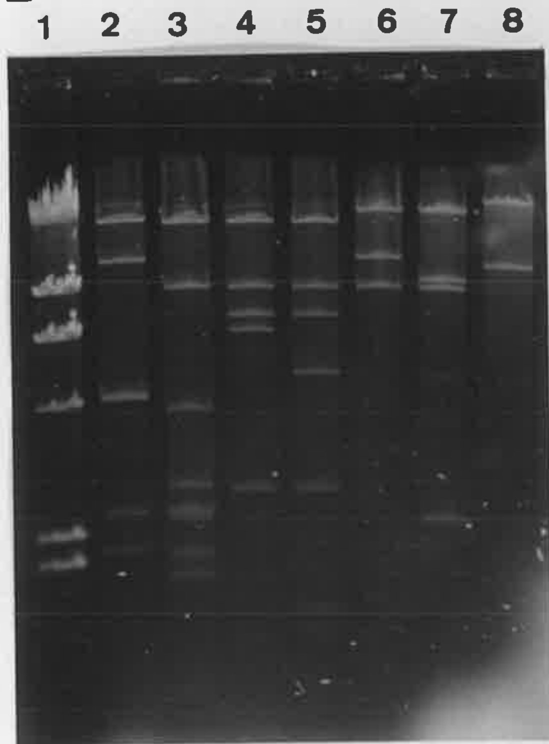
B) Agarose gel showing restriction enzyme products of clone 3-2 DNA upon digestion with 2) BamHI, 3) BamHI + EcoRI, 4) EcoRI, 5) EcoRI + XbaI, 6) XbaI, 7) XbaI + SacI and 8) SacI. Track 1 contains 0.5 μg of wild type lambda DNA restricted with HindIII.

C) Hybridization of pTB19 and pTB20 (which encompass nucleotide 26,191 to 55,921 of the tobacco plastid genome (Shinozaki *et al*, 1986) to a Southern blot of the gel depicted in B).

A

BamHI	BamHI-EcoRI	EcoRI	EcoRI-XbaI	XbaI	SacI-XbaI	SacI
L+1	L	L	L	L+5	L+5	L+15
R+2.8	R	R	R	12.5	R+0.03	R+2.4
4.54	4.19	7.4	7.39	R+0.3	10.0	
4.46	2.77	6.6	5.07		2.47	
2.51	2.51	2.8	2.8			
2.10	2.43	0.6	1.55			
	2.05		0.6			
	1.84					
	1.06					
	0.34					
	0.26					

B



C

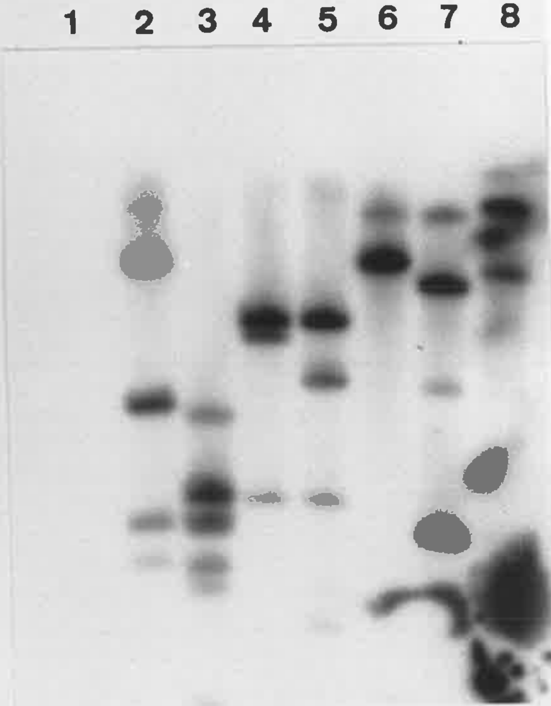
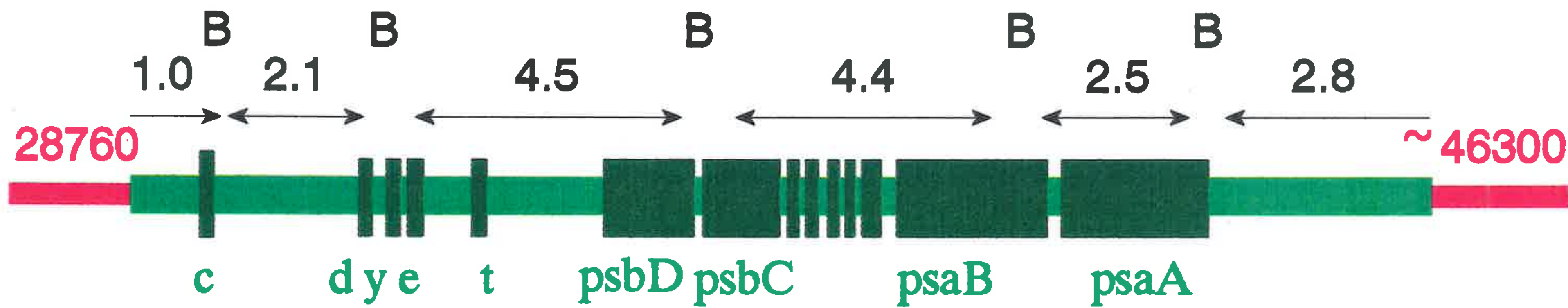


Figure 5.17: Restriction enzyme map of lambda clone 3-2.

The tobacco DNA insert of lambda clone 3-2 is represented by a light green line while flanking vector sequences are indicated by a thinner red line. BamHI, EcoRI, XbaI and SacI restriction sites are represented by a B, E, X and S respectively. Restriction fragments produced by digestion of clone 3-2 with these enzymes are indicated in kbp. Clone 3b shows homology to the tobacco plastid genome from approximately nucleotide 28,760 to 46,300 (Shinozaki *et al*, 1986) and plastid genes encoded by this clone are boxed in green and appropriately labelled. tRNA genes are labelled with a single lower case letter. The five genes flanked by *psb*C and *psa*B are not labelled and correspond to the following genes (from left to right): *trn*S, ORF62, *trn*G, *trn*fM and *rps*14.


B
B
B
B
B
1.0
2.1
4.5
4.4
2.5
2.8
28760
~46300
c
d y e
t
psbD psbC
psaB
psaA

E
E
E
E
E
2.8
0.6
6.6
7.4

X
X
5.0
12.5

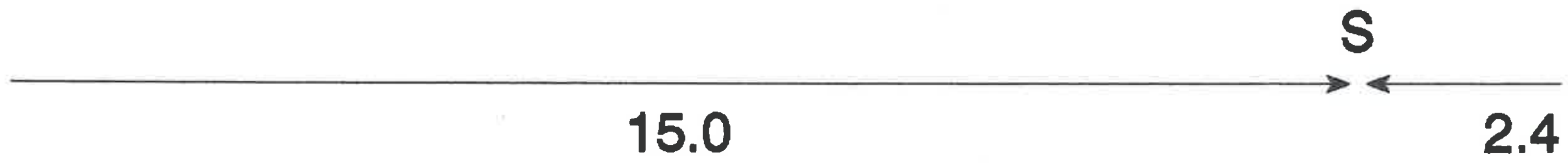

S
15.0
2.4

Figure 5.18: DNA sequence comparision between the tobacco plastid genome and the 0.6 kbp EcoRI restriction fragment of lambda clone 3-2.

The DNA sequence of the tobacco plastid genome is labelled "cp" and numbered according to the published sequence (Shinozaki *et al*, 1986). Nucleotides common to the plastid genome and the 0.6 kbp EcoRI restriction fragment of clone 3-2 (labelled 3-2) are indicated with "-" in the 3-2 sequence, while the positions of deleted nucleotides in the 3-2 sequence are marked with "■".

```
             31640     31650     31660     31670     31680
cp    TTATATTAAATTTTGTATTGGACAAGAAAGGAATTCCCCTTGTGTATGCGCGCCTCAAAA
3-2                                 ------------------------------

             31700     31710     31720     31730     31740
cp    AGGTATAGTACTCGATTCCATTACATGCATCGGGGGCAATCGAAAAAGCCAGCATTTCTT
3-2   ------------------------------------------------------------

             31760     31770     31780     31790     31800
cp    GGAATACTGACTATAATGCTACCAATAATCGTACTAATCCAACCGCATATGTCTTTCTCC
3-2   ------------------------------------------------------------

             31820     31830     31840     31850     31860
cp    TACCAAAAGGAAAGAAAAAAGAAATAAGGATTTCCCCTTTGCTTTGACAATGAAATTCTG
3-2   ------------------------------------------------------------

             31880     31890     31900     31910     31920
cp    CCCCCGGTCCCCTTCATAAAAAGGGAGAGATTTATTGATATATTTATTGGATCCATCGGG
3-2   ------------------------------------------------------------

             31940     31950     31960     31970     31980
cp    ACTGACGGGGCTCGAACCCGCAGCTTCCGCCTTGACAGGGCGGTGCTCTGACCAATTGAA
3-2   -----------------------------------------■------------------

             32000     32010     32020     32030     32040
cp    CTACAATCCCAGGGAAATACGGGATCTAGCAGAAAATTTGATTCTTTTTTATCTCCGGAT
3-2   ------------------------------------------------------------

             32060     32070     32080     32090     32100
cp    CGGGTATTTCTGAAGTACGAAGGGGGTTATATCATCTCATGGCGGATTGGCGAATTTTTG
3-2   ------------------A-----------------------------------------

             32120     32130     32140     32150     32160
cp    GGCCGAGCTGGATTTGAACCAGCGTAGACATATTGCCAACGAATTTACAGTCCGTCCCCA
3-2   ------------------------------------------------------------

             32180     32190     32200     32210     32220
cp    TTAACCGCTCGGGCATCGACCCAAGAAGAATCAATTTTAGACTTATTGGTAATCCATGAT
3-2   -----------------------G------------------------------------

             32240     32250     32260     32270     32280
cp    CAACTTCCTTTCGTAGTACCCTACCCCCAGGGGAATTCGAATCCCCGCTGCCTCCTTGAA
3-2   -----------------------■--------------
```

Figure 5.19: DNA sequence comparision between region 1 of lambda clone 3D and region 1 of lambda clone 3b.

The DNA sequence of region 1 of clone 3D (labelled 3D) is numbered according to the numbering of the 3D sequence in figure 5.4. Nucleotides shared by both 3D and region 1 of clone 3b (labelled 3b) are marked with a "-" in the 3b sequence, while the positions of deleted nucleotides in the 3b sequence are marked with "■". The first 10 nucleotides of the 3D sequence and the last 10 nucleotides of the 3b sequence represent the lambda EMBL4 cloning site.

```
                        10        20        30
3D           GAATTCCCCGGATCTCATATAGGCCTTTATGG-CCTGAC
3b  GCATGCCATGACGTTGTTA-----------------------C------

    40        50        60        70        80        90
3D  CTGTAAATGGACCTTTATGAGCTTCTAAAATATCTTTTAGTCCATGACCAATACCCCAGT
3b  ------------------------------------------------------------

    100       110       120       130       140       150
3D  TGGTCCTATACATGTGACCCGCTATCAGGAAAAGAATTGCAATAGCTAAATGGTGATGGG
3b  ------------------------------------------------------------

    160       170       180       190       200       210
3D  CAATATCAGTCAGCCACAGACCCCCAGTTACTGGATCTAATCCTCCACGAAAAGTAAGAA
3b  ------------------------------------------------------------

    220       230       240       250       260       270
3D  AGTCCGCATATTTTGACCAATTCAAGGTGAAAAATGGGGTTGCTCCCTCGGCAAAACTGG
3b  ----------------■-------------------------------------------

    280       290       300       310       320       330
3D  GATAAAGTTGAGCCAAAAGATCTCGATTCAAGATAAATTCATGAGGAAGTGGTATCTCTT
3b  ------------------------------------------------------------

    340       350       360
3D  TAGGATCTACTCCAGCGTTTAGA
3b  -------CGGGGAATTC
```

Figure 5.20: DNA sequence comparision between the 0.6 kbp EcoRI fragment of lambda clone 3-2 and region 2 of lambda clone 3b.

The DNA sequence of region 2 of clone 3b is labelled "3b" while that of the 0.6 kbp EcoRI fragment of clone 3-2 is labelled "3-2". Nucleotides common to both sequences are represented by "-" in the 3b sequence, while the positions of deleted nucleotides in the 3b sequence are represented by "■".

```
3-2   GAATTCCCCTTGTGTATGCGCGCCTCAAAAAGGTATAGTACTCGATTCCA
3b    --------------------------------------------------

3-2   TTACATGCATCGGGGGCAATCGAAAAAGCCAGCATTTCTTGGAATACTGA
3b    --------------------------------------------------

3-2   CTATAATGCTACCAATAATCGTACTAATCCAACCGCATATGTCTTTCTCC
3b    --------------------------------------------------

3-2   TACCAAAAGGAAAGAAAAAAGAAATAAGGATTTCCCCTTTGCTTTGACAA
3b    --------------------------------------------------

3-2   TGAAATTCTGCCCCCGGTCCCCTTCATAAAAAGGGAGAGATTTATTGATA
3b    --------------------------------------------------

3-2   TATTTATTGGATCCATCGGGACTGACGGGGCTCGAACCCGCAGCTTCCGC
3b    --------------------------------------------------

3-2   CTTGACAGGGC-GTGCTCTGACCAATTGAACTACAATCCCAGGGAAATAC
3b    -----------G--------------------------------------

3-2   GGGATCTAGCAGAAAATTTGATTCTTTTTTATCTCCGGATCGGGTATTTC
3b    --------------------------------------------------

3-2   TGAAGTACAAAGGGGGTTATATCATCTCATGGCGGATTGGCGAATTTTTG
3b    --------------------------------------------------

3-2   GGCCGAGCTGGATTTGAACCAGCGTAGACATATTGCCAACGAATTTACAG
3b    --------------------------------------------------

3-2   TCCGTCCCCATTAACCGCTCGGGCATCGACCCAGGAAGAATCAATTTTAG
3b    --------------------------------------------------

3-2   ACTTATTGGTAATCCATGATCAACTTCCTTTCGTAGTACCCTACCCC-AG
3b    -----------------------------------------------C--

3-2   GGGAATTC
3b    ■-------
```

Figure 5.21: Nucleotide sequence of the tobacco plastid genome region containing $tRNA^{Glu}$, $tRNA^{Tyr}$ and $tRNA^{Asp}$ genes.

Coding regions are boxed. "Pribnow box"-like and "-35 region "-like sequences are underlined. Large arrows indicate the transcription initiation (black) and termination sites (white), small arrows cleavage sites and horiziontal arrows direct repeat sequences and inverted repeat sequences.

Extracted directly from Ohme *et al*, 1985.

This sequence is the reverse complement of the tobacco plastid genome sequence published by Shinozaki *et al* (1986) and the sequence of clones 3b and 3-2 given in figure 5.15 and 5.18.

The "C" nucleotide which is marked with a cross and occurs two nucleotides before the start of the $tRNA^{Tyr}$ gene is published as a "T" (or "A" in the reverse complement) in the equivalent position (ie. nucleotide 32693) of the tobacco plastid genome sequence of Shinozaki *et al* (1986) (see figure 5.15 and 5.18). In clone 3b and 3-2 this nucleotide is also a "C" ("G" in reverse complement) (figure 5.15 and 5.18). Conversely the position of two nucleotides absent in this sequence, but found in the plastid sequence of Shinozaki *et al* (1986) (nucleotides 31680 and 31681) and the sequences of both clone 3b and 3-2 (figure 5.15 and 5.18) is marked with a large arrow and the deleted nucleotides are shown in brackets.

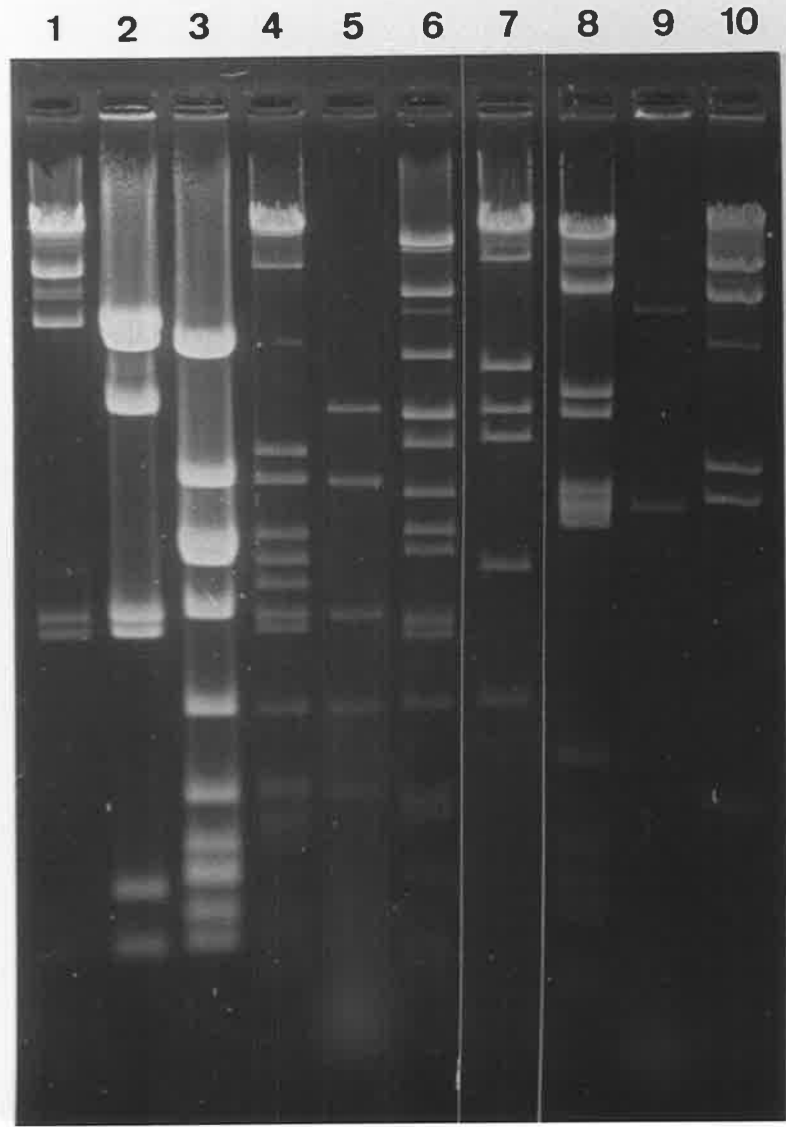
```
EcoRV                                                                                AvaII                    100
ATCTCGTCTTAACAATGAATCAAATGAAAGTGAAAGAAATAGAATTTCACACCTTTTTCCTTTTCTGACGGACCAATCATTCCCTGCAAAAATCCTACTC
                                                                                                              200
TTCCTTATTATTTCTATTTTTTGTATTTATTACTTTTTTTTTATTTATAAATAAAAAAAGAAAATTCTAAATACTAAATAATCTAAACTAAAATAATCGA
                                                                                                              300
AAGAAATTCAATTGAAATAATTCCAAAAAAAAAAAAAATACTACTACTAGATTTCTAATGGCGATTCTAATGAATAATTCATCAATGACGAATAAAAAAAT
                                              TaqI                         Sau3A                              400
TCTATGCAATTCTGAAAGGGGGAAAGATCCCTCGGATAGAATCATTCGATTATATTGACAATTTCAAAAAACTGATCATACTATGATCATAGTATGATGG
                                                          tRNA Glu                                            500
CGGTTGGTCAAGCAGGCCCCCATCGTCTAGTGGTTTAGGACATCTCTCTTTCAAGGAGGCAGCGGGGATTCGAATTCCCCTGGGGGTAGGGTACTACGAA
                                                                        TaqI EcoRI   tRNA Tyr                 600
AGGAAGTTGATCATGGATTACCAATAAGTCTAAAATTGATTCTTCCTGGGTCGATGCCCGAGCGGTTAATGGGGACGGACTGTAAATTCGTTGGCAATAT
           Sau3A                                                                                              700
GTCTACGCTGGTTCAAATCCAGCTCGGCCCAAAAATTCGCCAATCCGCCATGAGATGATATAACCCCCTTCGTACTTCAGAAATACCCGATCCGGAGATA
                         Sau3A                                            tRNA Asp                  TaqI      800
AAAAAGAATCAAATTTTCTGCTAGATCCCGTATTTCCCTGGGATTGTAGTTCAATTGGTCAGAGCACCGCCCTGTCAAGGCGGAAGCTGCGGGTTCGAGC
                                                                                                              900
CCCGTCAGTCCCGATGGATCCAATAAATATATCAATAAATCTCTCCCTTTTTATGAAGGGGACCGGGGGCAGAATTTCATTGTCAAAGCAAAGGGGAAAT
                BamHI (Sau3A)                                                                                 1000
CCTTATTTCTTTTTTCTTTCCTTTTGGTAGGAGAAAGACATATGCGGTTGGATTAGTACGATTATTGGTAGCATTATAGTCAGTATTCCAAGAAATGCTG
                                                                       CG
GCTTTTTCGATTGCCCCCGATGCATGTAATGGAATCGAGTACTATACCTTTTTGAGGCGCATACACAAGGGG
       TaqI                                                       HhaI          EcoRI
```

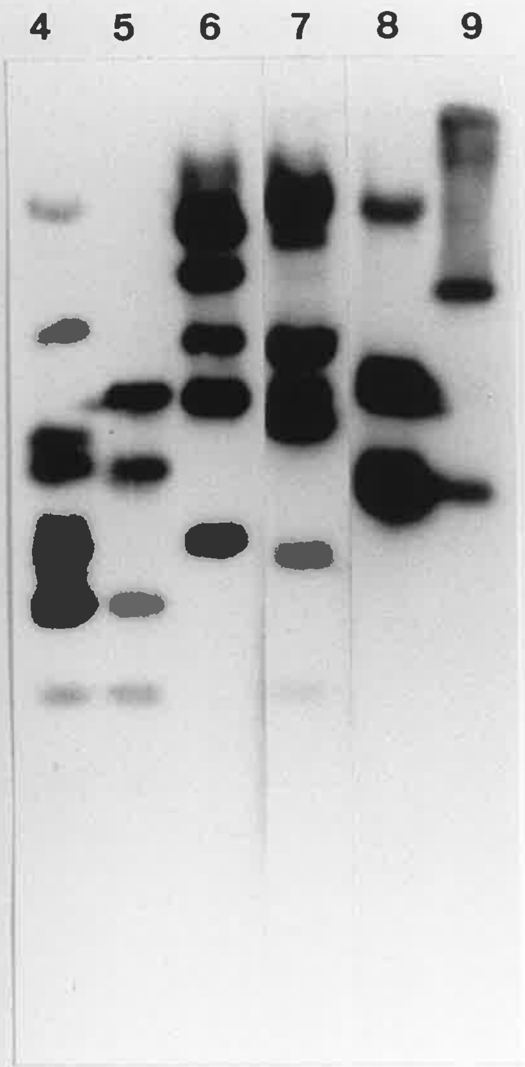
Figure 5.22: Restriction enzyme mapping of lambda clone 1H.

A) Agarose gel showing restriction enzyme products of lambda clone 1H DNA upon restriction with 1) BamHI, 4) EcoRI, 6) PstI, 7) XhoI and 8) BglII. Tracks 2 and 3 contain 2 μg of pTBa13 (which encompasses the tobacco plastid genome from nucleotide 83807 to 94562 (Shinozaki *et al*, 1986) restricted with BamHI and EcoRI respectively. The 5 kbp BamHI fragment of clone 1H which, is evident in track 1, was subcloned into the BamHI site of pUC19 and restricted with 5) EcoRI and 9) BglII. 0.5 μg of wild type lambda DNA restricted with HindIII was loaded into track 10 to serve as a molecular weight marker.

B) Hybridization of pTBa1, pTB13 and pTB18 (which covers the tobacco plastid genome from nucleotide 64101 to 104801 (Shinozaki *et al*, 1986) to a Southern blot of tracks 4-9 of the gel depicted in A).

C) Hybridization of pTBa1 which encompasses nucleotide 64101 to 83807 of the tobacco plastid genome (Shinozaki *et al*, 1986) to the same membrane used in B). Nucleotide 83807 occurs in the first exon of the *rps*16 gene (see figure 5.23).


1 2 3 4 5 6 7 8 9 10


4 5 6 7 8 9

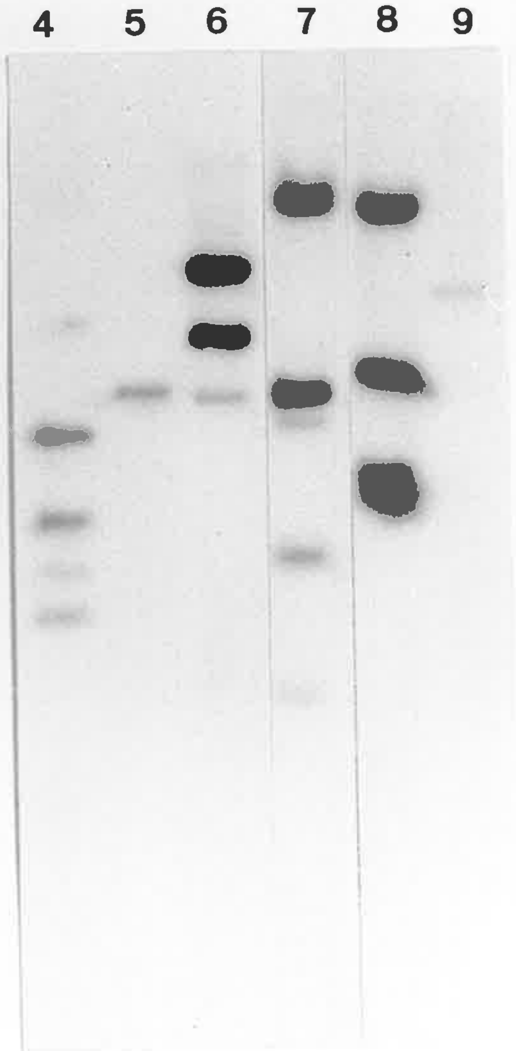
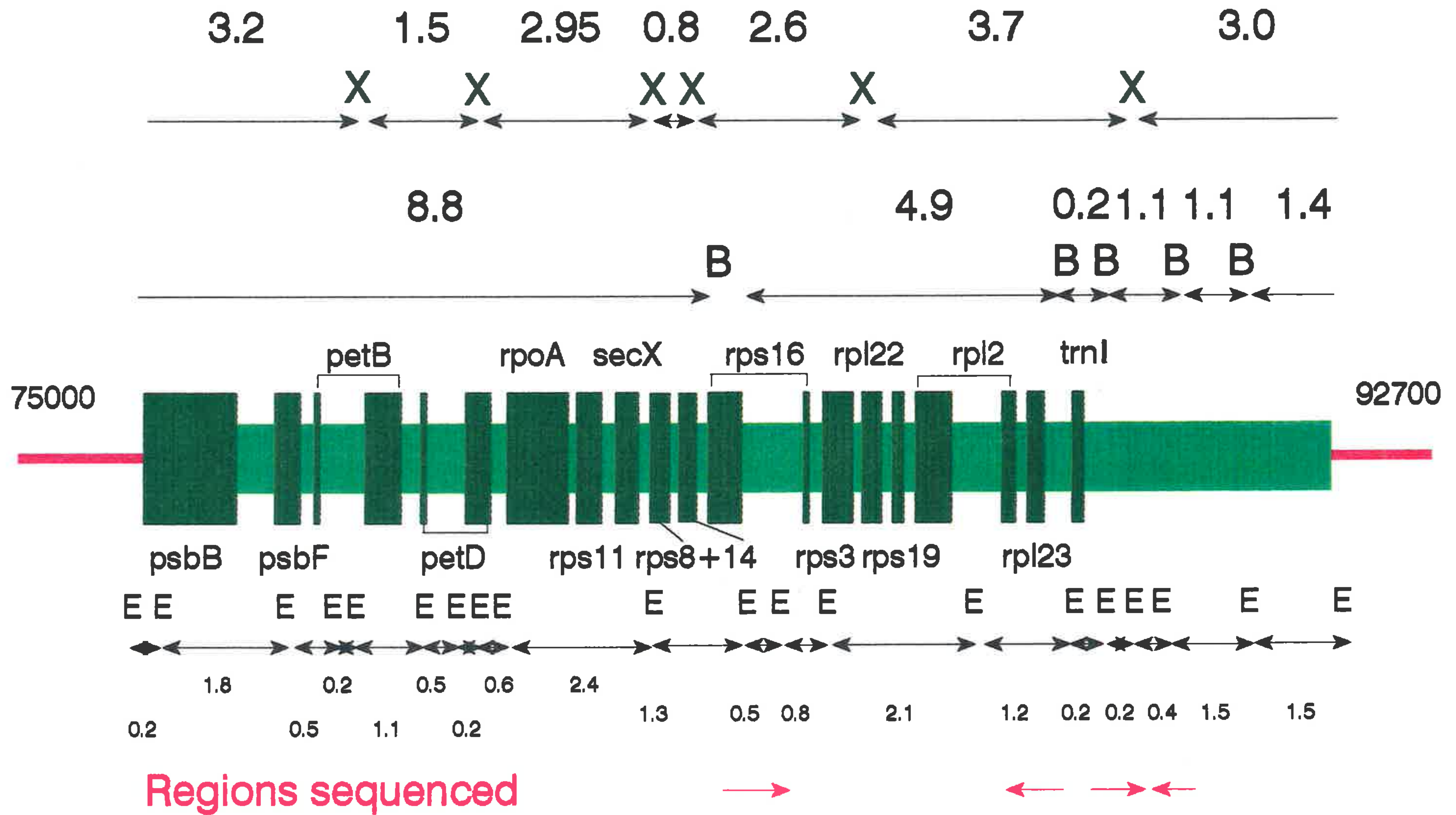

4 5 6 7 8 9

Figure 5.23: Restriction enzyme map of lambda clone 1H.

The tobacco DNA insert of lambda clone 1H is represented by a light green line, while flanking vector sequences are depicted as a thin red line. EcoRI, XhoI and BamHI restriction sites are marked with "E", "X" and "B" respectively. Restriction fragments produced by digestion of clone 1H with these enzymes are indicated in kbp. Clone 1H is homologous to the tobacco plastid genome from approximately nucleotide 75,000 to 92,700 (Shinozaki *et al*, 1986) and plastid genes encoded by this clone are boxed in dark green and labelled appropriately. Genes joined by brackets (eg. *rps*16, *pet*D and *rpl*2) contain introns and each exon of the gene is represented as an individual dark green box. Regions of clone 1H that have been sequenced are marked with red arrows.


3.2
1.5
2.95
0.8
2.6
3.7
3.0
X X X X X X
8.8
4.9
0.2
1.1
1.1
1.4
B B B B B
petB
rpoA
secX
rps16
rpl22
rpl2
trnI
75000
92700
psbB
psbF
petD
rps11
rps8+14
rps3
rps19
rpl23
E E E E E E E E E E E E E E E E E E E E E
1.8
0.2
0.5
0.6
2.4
0.2
0.5
1.1
0.2
1.3
0.5
0.8
2.1
1.2
0.2
0.2
0.4
1.5
1.5
Regions sequenced

CHAPTER 6

Lambda Clone 1M

Lambda clone 1M was isolated by screening the tobacco genomic library with the spinach *rbc*L gene and the tobacco plastid Bam8 fragment (Sugiura *et al*, 1986). Southern hybridization and restriction enzyme analysis localized the plastid homology of this clone to a single 1.5 kbp EcoRI fragment (figure 6.1). No homology between clone 1M and the mitochondrial genome was detected (chapter 5, figure 5.1). This restriction fragment in addition to the 1.8 kbp EcoRI restriction fragment that mapped immediately adjacent (figure 6.1) to this sequence were subcloned and into pUC19 and bidirectionally sequenced.

Clone 1M contained 1.3 kbp of DNA that showed 92% homology to the published sequence of the tobacco plastid genome, from nucleotide 90279 to 91621 which encodes part of ORF 581 and ORF 1708. (Shinozaki *et al,* 1986) (figure 6.2). This region of the plastid genome is not part of either probe but shows multiple short tracts (ie 30 bp) of homology to the probes used to isolate the clone. Given this limited homology it was fortunate that clone 1M was identified in the genomic library by these probes.

The remaining 2 kbp of this clone that was sequenced showed no homology to either the tobacco plastid genome or any other sequence on the EMBL, GenBank, Vecbase and NBRF databases. The 1.3 kbp region of plastid DNA homology differed from the tobacco plastid genome (Shinozaki *et al,* 1986) by 99 base substitutions, six insertions of 1 or 2 nucleotide pairs and five deletions ranging in size from 1 to 28 nucleotide pairs (figure 6.2). Assuming a rate of sequence divergence similar to that of mammalian pseudogenes [5×10^{-9} changes per site per year (Li *et al*, 1985; Dennis *et al*, 1988)] this region of the plastid genome is estimated to have transferred to the nucleus approximately 8 million years ago. This plastid integrant is approximately 10 times older than the promiscuous plastid sequence contained in clone 3D (chapter 5). A continual transfer of plastid DNA to the nuclear genome must therefore be occurring to

explain this result or alternatively these sequences must have integrated into regions of the nuclear genome undergoing very different rates of sequence evolution.

The plastid homology of clone 1M terminates at the lambda cloning site and so the absolute size of this plastid integrant is unknown. Only a single junction of nuclear and plastid derived DNA was contained in this clone, so little information regarding the transfer of this sequence to the nucleus can be inferred. However at the exact nuclear/plastid DNA junction is the sequence GCAAAACG (figure 6.2). Mirror image sequences of this sort have been suggested to arrest DNA synthesis (Weaver and DePamphilis, 1984) and may therefore be involved in plastid DNA insertion. Clone 1M is homologous to a large (7.7 kbp) intergenic region of the tobacco plastid genome, which is thought not to be transcribed within the plastid. Transfer of this sequence to the nuclear genome is therefore unlikely to involve an RNA intermediate, as postulated for some other promiscuous sequences (Schuster and Brennicke, 1987; Nugent and Palmer, 1991).

6.1: **Genomic location of clone 1M**

Hybridization of the 1.8 kbp EcoRI fragment of clone 1M, that is nuclear DNA flanking the plastid DNA insert, to total tobacco DNA restricted with either EcoRI or XbaI identified a broad spectrum of bands superimposed upon a background smear (figure 6.3, tracks 1, 2, 3). Raising the stringency of washing of this filter did not reduce the background appreciably (not shown). This hybridization pattern is consistent with that observed for repetitive sequences that are dispersed throughout the nuclear genome. Dotblot quantification gave an estimate of 400 equivalent copies of this repeated sequence per 1c tobacco nucleus. Dispersed repeats are believed to be derived from transposable elements (Flavell, 1988; John and Miklos, 1988) and it is significant that a sequence representing a transposition from the plastid genome should be associated with a nuclear sequence that potentially represents a transposition relic. No internal repeats could be identified within this sequence.

It is to be expected that this plastid sequence is associated with a repetitive DNA element given that repeated sequences form the bulk of many plant nuclear genomes. These regions of the nuclear genome are generally transcriptionally inactive and so integration of foreign sequences into this genomic location would be minimally detrimental to the plant. A majority of these repeated sequences were not restricted by HpaII (figure 6.3, tracks 4 and 5) demonstrating that they are heavily methylated, consistent with the hypothesis that they are located in transcriptionally inactive regions of the genome. These observations are also consistent with the apparent absence of plastid sequences within the nuclear genome of *Arabidopsis thaliana* (chapter 3, figure 3.13) which is an atypical plant species in that it possesses a small nuclear genome and little repetitive DNA (Meyerowitz and Pruitt, 1985).

6.2: **Deletions in lambda clones 1M and 3D**

Certain DNA sequences are known to undergo deletion mutations at much higher frequencies than others. Often associated with these deletion hotspots are DNA repeat motifs that frequently delineate the boundaries of the deletion (Albertini *et al*, 1982; Trinh and Sinden, 1991; Schaaper *et al*, 1986; Drake *et al*, 1983; Ripley, 1982). These repeat motifs include both direct repeats and palindrome sequences (Albertini *et al*, 1982; Ripley, 1982). The size of these sequence homologies can be as small as 5 bp, however there is a strong correlation between the frequency of deletion formation and the length of homology between the repeat sequences (Albertini *et al*, 1982).

When the double stranded DNA helix becomes unwound (eg. during replication) direct repeats can form transient, intrastrand DNA secondary structures that may cause deletion of the intervening sequence by DNA replication slippage. Secondary structures formed by palindromes can cause both deletions and frame shift mutations (Drake *et al*, 1983; Ripley, 1982). It has been shown in *E. coli* that replication-dependent deletion between direct repeats may occur preferentially in the lagging strand of the replication fork (Trinh and Sinden, 1991).

Many of the plastid DNA deletions identified in clone 1M and 3D have repeat motifs potentially capable of secondary structures analogous to those observed (Trinh and Sinden, 1991; Schaaper *et al*, 1986; Drake *et al*, 1983). [Assuming that these sequence differences are the result of DNA deletions in the nuclear sequences rather than DNA insertions in the plastid genome, subsequent to the transfer of these sequences from the plastid to the nucleus]. Within region 4 of clone 3D (chapter 5, figures 5.3 and 5.8) is a 41 bp deletion which is a region in the plastid genome that is flanked by two 11 bp direct repeats, one of which is lost upon formation of the deletion (figure 6.4a). During DNA replication of this region, misalignment of these repeats may form a single stranded DNA loop that is subsequently lost on one DNA strand (figure 6.4a). Similar mechanisms could explain the deletion of six nucleotide pairs (figure 6.4b) and the addition of two nucleotide pairs (figure 6.4c), in region 1 of this clone (chapter 5, figure 5.3). Also associated with a 12 bp deletion in region 4 of clone 3D are two overlapping 11 bp direct repeats present in the plastid genome (figure 6.4d).

Other processes such as homologous or illegitimate recombination or DNA repair activities could also offer explanations for these deletions (Trinh and Sinden, 1991). However the deletions present in these clones are relatively small and therefore unlikely to be the result of a recombination event. The probability of slipped mispairing between short homologies should increase as the distance between the homologies decreases (Albertini *et al*, 1983).

Clone 1M contains two deletions of 7 and 27 nucleotide pairs that are each surrounded by inverted repeats in the plastid genome from which they were derived (figure 6.5a and b). These sequences are potentially capable of forming palindrome mediated hairpin loops (Drake *et al*, 1983; Nag and Petes, 1991; Ripley, 1982; Schaaper *et al*, 1986). Imprecise enzymatic excision of these hairpin configurations may account for the deletions present in clone 1M (figure 6.5, a and b). Also associated with clone 1M are a further 2 deletions that are regions of direct repeats within the tobacco plastid genome (figure 6.5c).

A majority of studies examining DNA replication slippage and hairpin loop excision or frameshift mutation have been carried out in procaryotes (Schaaper *et al*, 1986; Trinh and Sinden, 1991; Albertini *et al*, 1982, Weaver and DePamphilis, 1984) or lower eucaryotes (Ripley, 1982; Nags and Petes, 1991) which are obviously the organisms of choice given the large number of individuals that must be screened to identify one of these rare mutation events. Promiscuous plastid sequences may be useful tools for examining nuclear mutational events in higher plants for several reasons. These sequences appear common to most plants so most species have a spectrum of plastid sequences that have been subjected to nuclear mutational events for variable lengths of time. The age of these sequences within the nuclear genome ensures that many will contain some form of rearrangement. The plastid genome has a very slow rate of sequence evolution so most changes present within a promiscuous plastid sequence will be the results of nuclear events. Many plastid genes have already been sequenced and so the DNA sequence prior to nuclear influence is essentially known, and suitable DNA probes for isolation of these sequences are readily available.

6.3: **Summary**

Clone 1M contains a 1.3 kbp region that was homologous to the tobacco plastid genome from nucleotide 90279 to 91612 (Shinozaki *et al*, 1986). This region of plastid DNA homology differed from the tobacco plastid genome by 99 base substitutions, six insertions and five deletions. The plastid integrant was located in an extensively methylated region of the nuclear genome that contained a member of a dispersed repeat family. Many of the plastid sequence deletions present in clone 1M and 3D represent regions of the plastid genome that either contain, or are flanked by, repeat motifs. DNA replication slippage or hairpin loop excision may account for these deletions. The region of plastid sequence homology of clone 1M represented a transposition event from the plastid to the nucleus that took place approximately 8 million years ago. As the plastid DNA insertion contained in clone 1M is 10 times older than that contained in 3D, transposition of plastid sequences to the tobacco nuclear genome may be an

ongoing evolutionary process. Alternatively the different amounts of sequence divergence between these two clones could be the result of their integration into different genomic locations that undergo different rates of sequence evolution.

Figure 6.1: Restriction enzyme map of lambda clone 1M.

The tobacco DNA insert of clone 1M is represented by a thick purple and green line with flanking vector sequences depicted as a thin red line. The position of EcoRI restriction sites are marked with an "E" and the size of EcoRI restriction fragments given in kbp. The region of plastid DNA homology of this clone is shaded in green, appropriately labelled and numbered according to the published plastid DNA sequence (Shinozaki *et al*, 1986). The remainder of clone 1M showed no homology to either the tobacco mitochondrial genome (chapter 5, figure 5.1) nor the tobacco plastid genome.

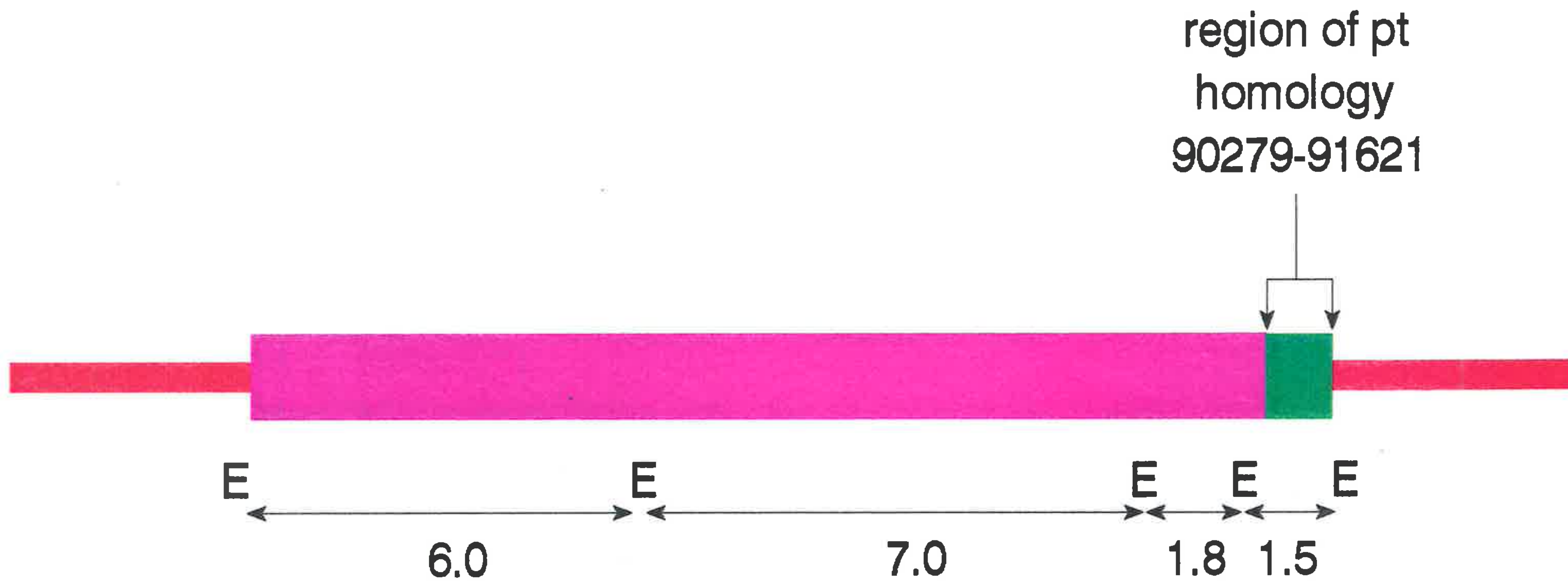

region of pt
homology
90279-91621
E
E
E
E
E
6.0
7.0
1.8
1.5

Figure 6.2: DNA sequence comparision between the tobacco plastid genome and the 1.5 and 1.8 kbp EcoRI restriction fragments of lambda clone 1M.

The sequence of lambda clone 1M is illustrated such that the first six nucleotides correspond to the EcoRI site of the 1.8 kbp EcoRI fragment closest to the left arm of the vector (see figure 6.1). Nucleotides common to the plastid genome and clone 1M are marked with a "-" in the plastid sequence (cp), whereas the position of inserted nucleotides in the 1M sequence is marked by a "■" in the corresponding plastid sequence. Numbering of the plastid sequence is according to the published sequence of the tobacco plastid genome (Shinozaki *et al*, 1986). EcoRI restriction sites are underlined. The last 10 nucleotides of the 1M sequence correspond to the lambda EMBL4 cloning site and consequently show no homology to the plastid genome.

1M GAATTCGAGATGAAAGATCTCGAAAAGACAAAATTATGTCTTGGTTTGCAAA

1M TTGAACATTTGGCAAGCGAGATTTTTGTTCATCAATCTGCCTACATAGAAAAGGTATTGA

1M AACGGTTTTATATGGATAGAGTACATCCATTAAGTACTATGATGATTGTTCGATCACTTG

1M ATGTGAATAAGGACCCGTTCCGACCTCAAGAAAAGAATGAAGAACTTCTTGGTCCTGAAG

1M TACAATATCTTAGTGCAATTGGTGCACTAATGTATCTTGCTAATACTACAAGGCCTGACA

1M TAACTTTTTCAGTTAATGTCTTAGCAAGATATAGATCTGCTCCTACAAGGATACATTAGA

1M ATAGAATTAAACACATATTGCAGTATCTAAAACGGACTACCGATATGGGCTTATTTTATG

1M GCAATGATTGGAGTCCCAATATTCTTGGTTGTGCTGATGCTGGGTATATGCTGGGTATTT

1M ATCTGACCCACACAAAGCTCGATCTCAAACAGGCTATGTGTTTACATGTGGAGGCACTGC

1M AATAACTTGGCGATCGACTAAGCAATCAATCGTGGCTACTTCATCTAATCATACTGAGAT

1M AATTGCTATTGATAAAGCAAGTCGAGAATGTGTATGGTTGAGGTCTATAATACACCTTAT

1M TCGAGACAAATGTGGTTTGAAGTGTGACAAACTACCCATAATTTTATATGAAGACAATGC

1M AGCATGCATAGCCCAATTGAAGGGAGGATTCATAAAAGGAGATAGGACAAAGCACATTTC

1M ACCAAAGCTATTTTTCACAAATGATCTTCAAAAGAATGGTGATATCAATGTGCAACAAAT

1M CCGTTCAAGTGATAATATGGTTGATTTGTTCACCAAATCTCTACCGACGTCCACCTCAAG

1M AAACTAGTGTACAAGATTGGGATGTGAAGGCTCAAGGATGTGAATTGATGCTCTCATCAG

1M GGGGAGTTAATATGCGGTGTACTCTTTTTCCCTTATAAGTTTTTGTCCCACTGGGTTTTC

1M CTTGCAAGGTTTTTAATGAGGCAACCAAAAGGTGTATTTCTAAACATGTGTACTCTTTTT

1M CCTTCACTAGAATTTGTTTTTCAATAAGGTTTTAACGAGGCACATTATCTATGGACATCC

1M AAGGGGGAGTGTTATAAGAAAAATCAAATTATGGTGGATGTCTACCTTCTTCCATGATCT

1M TCCTCTCAAATGCTTAATCACATATTCAATGACACATTTCTTCATTTTTCATGCCTATAT

1M AAAGGCATTGTAATAGATGGAAAGAATACACAATTGAAGAAGAAATAAAATCCCTTCCTT

1M CTCTCTATCTCCATTTCTTGTTAATGTTTTACTAAATTGCTTTTATTTCATAACATTTTG

1M GTTTATTTTAGTTTTTAGTTTTTTTTCTTCTTCTTTTTCTTTTTGTTTTTCTTTAAAAGA

1M TAAAATATAAAGATTAAAAAAAATCAACTATCTAATTTTTAGGCTAAGAAAGATATAGGT

1M AAACTAAATATTTACACTATAATTTAAATGACACCAAAAATATACTAAAGCTAAATTATT

1M TATTATAATTCTAATTAAGAGAAACCATATATTTTTGTAAATGTTTGCTTTAAATAGAAA

1M TAAAACTTGTACTCTAAGAATGTGAATATTTTTTGTAGTTTTTAATTTTCATAAATAAAA

1M CCTATTAAGAGTAAGATAAAAGTTTAAAATATTAAGATTAGACGAAAAAGAAATATTTAC

```
1M      ACTAAAGTAGTATAGAATTCAGGTGCGGTCAAAAATTAGTTGATCACAAACTCTTCCAAC

1M      TTAAAGCTTCCAAATTGAGAATTATTCCTCCGAATCAACTCCAAACATCCCAAAAATCAA

1M      AATCAACCATACATGCAAGTCATAATGCATAAAGTGAAGCTACTCAAGGTCTCAAACTGC

1M      CGAACGACATCCTAGAGCTCAAAATGACCAGTCGGGTCGTTACGCAAAACGTGAGAAGCA
cp      -CCGAA-TGC-T-TCT-GAT-TTCCTCAAT---CC-G-TA-T-A-GG-------------
             90240     90250     90260     90270     90280     90290

1M      GATGATTAATCATCTATTTCCGAAAGAAATCGAAGAATTTCTTGGGAATCCTACAAGATC
cp      ---------------G------G-------------------------------------
             90300     90310     90320     90330     90340     90350

1M      CGATCGTTGTTTTTTTCTCTGATAGATGGTCAGAACTTCATTTGGGTTCGAATCCTACTG
cp      --T-----■■-------------------------------C------------------
              90360     90370     90380     90390     90400     90410

1M      AGAGGTCCACCAAAGA-CAGAAATTGTTGAACAAACAAACAAGATCTTTCTTTTGTCCAT
cp      ----------T-GG--T---------------G----■---------------------C-
              90420     90430     90440      90450     90460

1M      -------GGTTGGAACATAAAGAAATGGTTAATATATTCAAGATAATTACGTATTTACAA
cp      TCCAAGC-A-C----A--------------------------------------------
    90470     90480     90490     90500     90510     90520


1M      AATACCGTCTCAAT---TCCTATTTCATAAGATCTGAAACGTGATATGGTTTTGAAGTAT
cp      --------------TCA------------C-----C-GG-T------------CC----■--
    90530     90540     90550     90560     90570     90580

1M      GAACTGGATATAGACAGTTCTAACAAGATTTCATTCTTGAACAAACATCTTATTTTTAAA
cp      ----C------G--------C--T---------------------A---■C------TG-
     90590     90600     90610     90620     90630      90640

1M      TTTATTTTATCCGTTCCATGACCACAACAAGGAAGGATACACGATACACTACGATTTTGA
cp      -------C---TA----------GG----G--G----------T----------------
      90650     90660     90670     90680     90690     90700

1M      ATCAAAAGAGAGATTTCAAGAAATGACAGATCTATTCACTCTATCAATAACTGAGCCTCA
cp      ----G--------------------G--------------------------C-----GG-
      90710     90720     90730     90740     90750     90760

1M      CCTGGTGTATCATAAGGGATTTGCCTTTTCTATTGATTCCTGCGGATAGGATCAAAAACA
cp      T-------------------------------------------T-------------
      90770     90780     90790     90800     90810     90820
```

```
1M    ATTCTTGAATGAGGCCACGGATGAATTGAAAAAGAAATTTTTATTGGTTCTACCTCTTAT
cp    -----------------G--------C-----------C-----------------C---
    90830     90840     90850     90860     90870     90880

1M    TTTTTATGAAGAGAATGAATCTTTTTCTCGAAGGATTACAAAAAAATGGGTCTGGATCCC
cp    ------------------------------------C-G-------------C-----T-
    90890     90900     90910     90920     90930     90940

1M    CTGCGAGAATGCTTTGGAAGATCCAAAACCAAAAATAGTGGTATTTGCTAGCAATAACAT
cp    -----G-----A------------------------------------------C-----
    90950     90960     90970     90980     90990     91000

1M    AATAGAGGCAGTCACTCAATATAGATTGATCCAAAATCTGACTCAAATTGAATATAGTAC
cp    ---G----------------------------G--------T------CC----------
    91010     91020     91030     91040     91050     91060

1M    CTCTGGGTACATAAGAAATGTATTGAATCGATTCTTTTTAATAAATAGATCTGTACACAA
cp    --A---------------------------------------G--------C-AT-G---
    91070     91080     91090     91100     91110     91120

1M    ACTCGAATATGGAACTCAAAGGGATCAAATAGGAAAGGATACTCTGAATCATAGAACTAT
cp    CT------------T---------------------------------------------
    91130     91140     91150     91160     91170     91180

1M    AATGAAATATATGATCAACCAATATTTATCGAATTTTAAAAAGAGTTTGAAGAAATTTTT
cp    -----------C------------------------G---------CA--------GG--
    91190     91200     91210     91220     91230     91240

1M    C----------------------------TAGAGATCGATGAATCAGTATCCTGATGCAT
cp    -GAGCCTCTTATTTTGATTTCTCGAACCG■-------C-------G-G------------
    91250     91260     91270      91280     91290     91300

1M    ATAGATACAAATGGTCCAATGGAAGCAAGAATTTCTAGAGACATTTAGAACAGTCCATTT
cp    ----------------------G------------C--GA------G---------G---
     91310     91320     91330     91340     91350     91360

1M    CGGAGCAGAAGAGCCATTTTCAAGTAGTGTTTGATTGATTAAATATTAATCAATATTCGA
cp    ---------------G---------------C---C-----CG-----------------
     91370     91380     91390     91400     91410     91420

1M    TTGATTGGTCTGAGGTTATCCACAAAAAAAAGATTTGTCTAAGCCACTTCATTTCGTTTT
cp    --------------------G--■■-------------------------G----T----
     91430     91440       91450     91460     91470     91480
```

```
1M      GTCCAAGTCA---TTATTTTTGTCCAAGTTGCTTTTCTTTTTGTCTAGCTCACTTCCTTT
cp      ----------CTTC-T-------------------------------A------------
         91490     91500     91510     91520     91530     91540

1M      TTTCTGTGTGAGTTTCGGAAATATCCCTACTCGTCGGTCCGAGATCTTTATCTATGAATT
cp      ---------------------------C-T--A-A------------AC-----------
         91550     91560     91570     91580     91590     91600

1M      GAAAGGTCGGAATGATCCGGGGAATTC

Tobcpc  --------C--------AACTCTGCAATCAGTTGTTAGAATCAATAGGTCTTCAAATTGT
         91610     91620     91630     91640     91650     91660
```

Figure 6.3: The 1.8 kbp EcoRI restriction fragment of clone 1M is homologous to a moderately repetitive, dispersed nuclear repeat.

Hybridization of the 1.8 kbp EcoRI restriction fragment of lambda clone 1M (see figure 6.1) to total tobacco DNA. Each track contained 5 ug of DNA restricted with: 1) XbaI, 2) XbaI + EcoRI, 3) EcoRI, 4) HpaII + EcoRI and 5) HpaII.

Arrows indicate molecular weights of 9.4, 6.6, 4.4, 2.3, and 2.0 kbp.

1 2 3 4 5

Figure 6.4: Potential DNA replication intermediates resulting in the formation of deletions and insertions in lambda clone 3D by replication slippage.

A) Displacement of the DNA template strand by complementary base pairing between opposite DNA strands by direct repeats during DNA replication. The net result is a 41 bp deletion present in this region of plastid DNA sequence homology of clone 3D. The nucleotides deleted from the original plastid sequence are shown in green while direct repeats are underlined with an arrow. A dotted line indicates the direction of DNA replication. Numbering of the plastid sequence corresponds to that of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986).

B) Displacement of the DNA template due to incorrect complementary base pairing between the template and synthesizing DNA strand. A 6 bp region of the plastid genome (shown in green) is subsequently deleted during replication of the nuclear sequence encoded by clone 3D in one DNA strand. A dotted line indicates the direction of DNA replication. Numbering of the plastid genome corresponds to that of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986).

C) Insertion of 2 bps into a region of sequence of clone 3D by DNA replication slippage. The inserted nucleotides in the 3D sequence are shown in green and the numbering of the plastid genome corresponds to that of the published sequence (Shinozaki *et al*, 1986).

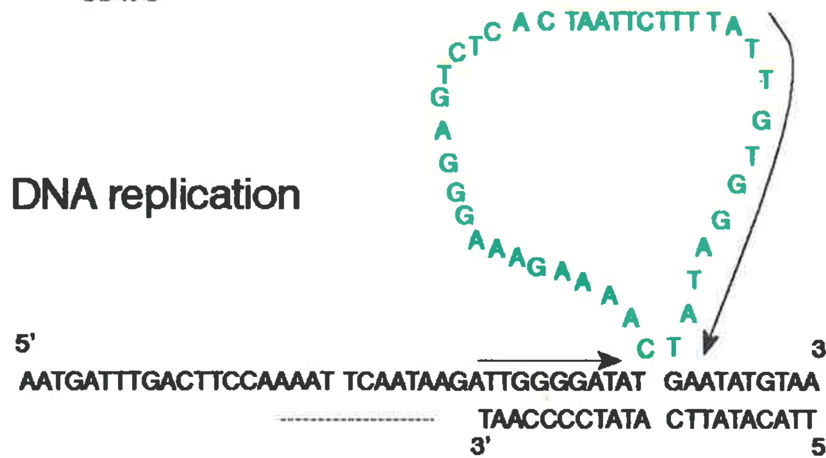
D) 11 bp direct repeats present in the plastid genome (underlined with green arrows) are associated with a 12 bp deletion present in clone 3D. Numbering of the plastid genome corresponds to that of the published sequence (Shinozaki *et al*, 1986).

A

Plastid genome

5' TCAATAAGATTGGGGATAT CAAAAGAAAGGGAGTCTCACTAATTCTTTTATTGTGGATATGAATATGTAA 3'

53470 53530



Clone 3D

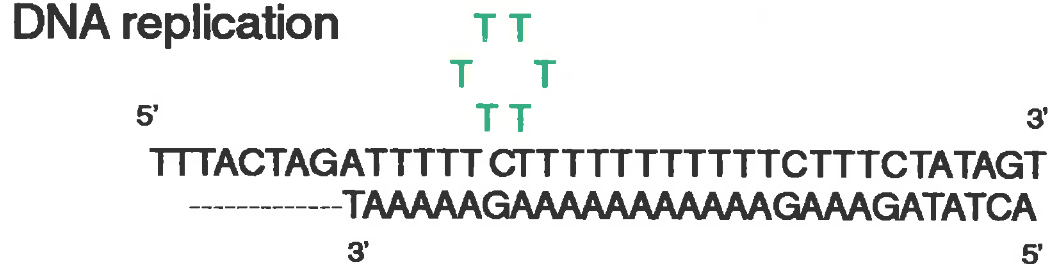
5' AATGATTTGACTTCCAAAATTCAATAAGATTGGGGATATGAATATGTAA 3'

B

Plastid genome

5' TTTACTAGATTTTTCTTTTTT TTTTTTTTTTTTCTTTCTATAGT 3'

45110 45150



Clone 3D

5' TTTACTAGATTTTTCTTTTTTTTTTTTCTTTCTATAGT 3'

C

Plastid genome

5' 3'

CCCCGCTATATATATTTA

43890 43900

DNA replication

TA

5' 3'

CCCCGC TATATATAT---------

GGGGCGATATATATAAAT

5'

Clone 3D

5' 3'

CCCCGCTATATATATATATTTA

D

Plastid genome

TATAGAATTTTTTGGATTTTGCATTTTTATAGGG

53720 53750

Clone 3D

TATACAATTTT — — — — TTTTTATAGGG

Figure 6.5: Some deletions in clone 1M are associated with regions of the plastid genome that contain palindromes which are potentially capable of forming hairpin loops.

A) Flanking a 27 bp deletion in clone 1M is a region of plastid sequence that may form a 32 bp hairpin loop (inverted repeats are shown in green). Base substitutions required to form the final 1M sequence are shown in purple. Arrows identify breakpoints in the DNA sequence. The numbering of the plastid sequence corresponds to that of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986).

B) Associated with a 7 bp deletion present in clone 1M is a region of the plastid genome that may form a 39 bp hairpin loop. Complementary regions of the plastid sequence are shown in green and the numbering corresponds to that of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986). Arrows indicate the sites of DNA breakpoints and base substitutions required to form the 1M sequence are shown in purple.

C) Direct repeats in the plastid genome (underlined with green arrows) are associated with deletions present in clone 1M. Numbering of the plastid sequences is according to the numbering of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986). A base substitution in the 1M sequence is coloured in purple [C(ii)].

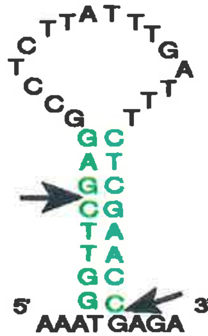
A

Plastid genome

5' AAATGGTTCGAGGCCTCTTATTTTGATTTCTCGAACCGAGA 3'

91250 91270

Rearrangement



Clone 1M

5' AAATTTTTCTAGA 3'

B

Plastid genome

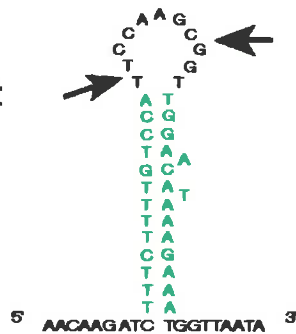
5' ATCTTTCTTTTGTCCCTTCCAAGCGATCGGAAAATAAAGAAA 3'

90460 90490

Base substitutions

5' ATCTTTCTTTTGTCCATTCCAAGCGGTTGGAACATAAAGAAA 3'

Rearrangement



Clone 1M

5' ATCTTTCTTTTGTCCATGGTTGGAACATAAA 3'

C(i)

Plastid genome

Clone 1M

GTCTCAAT— —TCCTATTTCATAAGAT

C(ii)

Plastid genome

Clone 1M

CCAAGTCA - - - CTTTTTTTGTCCAAGTTGCTTTTCTTTTTGTCTAG

CHAPTER 7

Transcription Of Promiscuous Plastid Sequences In The Plant Nuclear Genome

A plastid sequence that has integrated into the nuclear genome has a number of barriers preventing transcription and/or translation of the genes which it carries. Transcription and translation of plastid genes in the nucleus would be prevented by a requirement for different sequence motifs for the initiation of nuclear gene expression. However several scenarios could lead to expression of a promiscuous plastid sequence and presumably must occasionally occur to account for the endosymbiotic origin of nuclear encoded plastid proteins. If the invading sequence integrated into a nonessential nuclear gene it would be transcribed and possibly translated as part of a chimeric sequence. Alternatively a promiscuous sequence that integrates into an intron of a nuclear gene (eg. a region of the *psb*G gene into the third intron of the *cab*7 gene (Pichersky and Tanksley, 1988) would be transcribed (Piechulla *et al*, 1991), but subsequently spliced out of the transcript during RNA processing. A final possibility (although presumably very rare) is that a promiscuous plastid sequence may contain a gene promoter that can initiate transcription in both the plastid and the nucleus. For example the tobacco *psb*A gene promoter has been demonstrated as capable of initiating nuclear transcription (Cornelisson and Vandewiele, 1989).

Described in this chapter is a preliminary attempt to determine if promiscuous plastid sequences present within the nuclear genomes of both tobacco and beet give rise to any polyadenylated mRNA. A tobacco transcript with homology to plastid DNA was identified only in a polyadenylated RNA enriched fraction suggesting that it was transcribed within the nucleus. Subsequent attempts to clone this specific transcript were unsuccessful. A large cDNA clone with homology to the same region of the plastid genome was isolated but was of undetermined origin.

7.1: **Northern Blot Analysis**

Northern blots of polyadenylated mRNA from both tobacco and beet were hybridized with the spinach plastid *rbc*L gene and the spinach plastid 7.7 kbp PstI DNA fragment. As neither plastid nor mitochondrial transcripts are polyadenylated any transcripts with plastid homology that are enriched in a polyadenylated RNA fraction, but are reduced or absent in total RNA with polyadenylated RNA depleted, are presumably of nuclear origin. Figure 7.1b track 1 shows that four transcripts with homology to the spinach plastid 7.7 kbp PstI DNA fragment are present in a tobacco polyadenylated RNA enriched fraction. Three of these transcripts are also found to be abundant in both total RNA (figure 7.1b and c, track 3) and polyadenylated RNA depleted samples (figure 7.1b and c, track 2) and so are probably due to plastid RNA contamination of the polyadenylated RNA fraction. However the fourth transcript (marked with an arrow in figure 7.1b, track 1) is only present in the polyadenylated RNA fraction. This transcript of approximately 2.5 kbp is not identifiable in total RNA because of the intense hybridization signal from the abundant *bona fide* plastid transcripts. Therefore this comparatively low abundance transcript with homology to plastid DNA is likely to be a polyadenylated mRNA of nuclear origin.

No polyadenylated transcript with homology to the 7.7 kbp PstI DNA probe was detected in RNA from beet (figure 7.1a, track 1) and neither beet nor tobacco produce a detectable polyadenylated transcript with homology to the spinach *rbc*L gene (results not shown).

7.2: **Tobacco cDNA Library**

A tobacco cDNA library was constructed by the method of Gubler and Hoffman (1983) in lambda gt10 in an attempt to clone a sequence complementary to this polyadenylated tobacco mRNA, with homology to the spinach plastid 7.7 kbp PstI DNA fragment. The total library contained 1×10^6 pfu of which 2×10^5 were recombinants as determined by selective growth on *E.coli* C600 *Hfl*. Screening of this library with the spinach plastid 7.7 kbp PstI DNA probe identified a single positive clone. Further

screening of the library with plasmid clones encompassing the entire tobacco plastid genome identified no more cDNA clones with plastid DNA homology.

7.3: Tobacco cDNA Clone With Homology To The 7.7 kbp PstI DNA Fragment

The single positive lambda clone from the cDNA library was a very slow growing bacteriophage and produced a plaque approximately one fifth the size of those produced by most other recombinant phage in the library. Consequently six plating steps were required for its purification.

Figure 7.2b shows that this lambda clone is homologous to the spinach plastid 7.7 kbp PstI DNA probe. However it does not have a restriction map consistent with that of a recombinant lambda gt10 bacteriophage. EcoRI restriction of a recombinant lambda gt10 phage genome should produce restriction fragments of 32.7 and 10.6 kbp, while BamHI restriction should produce 5.5 and 9.8 kbp fragments and a 9.4 kbp fragment should be produced by HindIII restriction. This inconsistency may be the result of extensive rearrangement of the phage during repeated passage in *E. coli* C600 *Hfl*, a *recA*$^{+}$ host. Bacteriophage containing rearranged genomes often produce small, slow growing plaques. Four separate single plaques were isolated and DNA prepared and restricted from each (figure 7.2c). An identical restriction pattern was observed for each indicating that if rearrangements had occurred at some time during cloning and purification, the bacteriophage was now a stable if corrupted clone.

All the homology of this lambda clone to the spinach plastid 7.7 kbp PstI DNA fragment was contained in a single 4.4 kbp HindIII fragment (figure 7.2b, track 3). This fragment was subcloned into pUC19 and unidirectionally sequenced in several regions (figure 7.3). A total of 1787 bp of sequence was obtained that was continuously homologous to the tobacco plastid genome. From this sequence data, in addition to the size and restriction map of this fragment, it was concluded that this 4.4 kbp HindIII fragment represents a stretch of tobacco plastid DNA encompassing nucleotide 36,392-40,996 of the published plastid genome sequence (Shinozaki *et al*, 1986). This cDNA

clone is therefore significantly larger than the 2.5 kbp polyadenylated transcript identified by Northern blot analysis.

7.4: **Origin Of This cDNA Clone**

Although highly homologous to the tobacco plastid genome this sequence was divergent, differing by three deletions, four insertions and one base substitution. As the sequence was obtained from one DNA strand only, this level of divergence may be reduced by bidirectional sequencing. However certain completely unambiguous sequence changes are present in this cDNA clone that destroy ORFs in both the *psa*B and *psb*C genes. Furthermore these sequence changes destroy all three potential ORFs of the mRNA and so further sequence changes either upstream or downstream of these regions could not restore the *psa*B or *psb*C gene products.

It is possible for cDNA libraries to be contaminated with clones derived from nonpolyadenylated transcripts, by a process of self priming during first strand synthesis of cDNA by reverse transcription. As plastid transcripts frequently contain 3' inverted repeats they may be even more susceptible to this event (Stern and Gruissen, 1987; Gruissen, 1989). However this cDNA clone is very unlikely to be the product of a self primed plastid RNA molecule as such an RNA molecule would encode for nontranslatable versions of the *psb*C and *psa*B genes. Both of these genes encode thylakoid membrane proteins that are essential for plastid photosynthesis. In addition the *psa*B and *psb*C genes are transcribed from opposite DNA strands of the tobacco plastid genome and are therefore not transcribed as a single transcription unit (figure 7.4) (Yao et al,1989). Although this cDNA clone is reasonably large, it is noteworthy that the tobacco *psb*D-*psb*C gene cluster can be transcribed as a single 4.4 kbp transcript (figure 7.4) (Yao *et al*, 1989) and the spinach *psa*A gene is transcribed as part of a 6 kbp mRNA (Westhoff *et al*, 1983).

If this cDNA is not of plastid origin it can only be of either mitochondrial or nuclear origin. This region of the tobacco plastid genome has been detected in both the nuclear and mitochondrial genome (chapter 5, figures 5.12 and 5.1). A mitochondrial

origin of this clone would explain the apparent absence of this transcript in a Northern blot of a polyadenylated RNA fraction in addition to both the sequence divergence and cotranscription of the *psa*B and *psb*C genes. However as mitochondrial transcripts are not polyadenylated and this sequence was derived from oligo dT primed reverse transcription it is unlikely to be of mitochondrial origin. If of nuclear origin, this transcript could either be not polyadenylated and thus not present in polyadenylated RNA (and therefore very unlikely to be cloned) or polyadenylated but at a copy number too low to be detected by Northern blot analysis. Perhaps this clone represents an unprocessed version of the polyadenylated mRNA with homology to the 7.7 kbp PstI DNA fragment (figure 7.1b, track 1). A nuclear origin would also explain the sequence divergence and cotranscription of the *psa*B and *psb*C genes. Either a nuclear or mitochondrial origin of this cDNA clone represents transcription of a promiscuous plastid sequence.

Alternatively the insert of this lambda clone may have arisen from a DNA fragment contaminating the tobacco RNA preparation. However this DNA fragment would have to be of either nuclear or mitochondrial origin as the genes encoded by this insert are non-functional. As promiscuous plastid sequences in the nucleus account for only approximately 1% of chloroplast sequence homology present in total tobacco DNA (chapter 4) it would unlikely for a nuclear sequence to be cloned *in lieu* of true plastid DNA. This region of the tobacco plastid genome may also be integrated into the mitochondrial genome as evident from the homology between the 6.6 and 3.9 kbp EcoRI restriction fragments of clone 3b and the mitochondrial genome (chapter 5, figure 5.1). Therefore this cloned fragment may have arisen from mitochondrial DNA contamination of the tobacco RNA preparation.

7.5: **Summary**

These results suggest that in tobacco there may be expression of plastid sequences that have integrated into the nuclear genome. The identification of a polyadenylated RNA with homology to the plastid genome suggests a nuclear origin for

this transcript. The relatively high copy number of plastid sequences integrated into the tobacco nuclear genome and the ability of some plastid promoters to initiate nuclear transcription (Cornelisson and Vandewiele, 1989) supports the notion of expression of promiscuous plastid sequences. The cDNA clone isolated in this study contained rearrangements and was therefore of indeterminate origin.

Figure 7.1: Identification of a polyadenylated transcript with plastid DNA homology.

A) Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to RNA isolated from beet green leaf tissue. RNA samples consisted of:

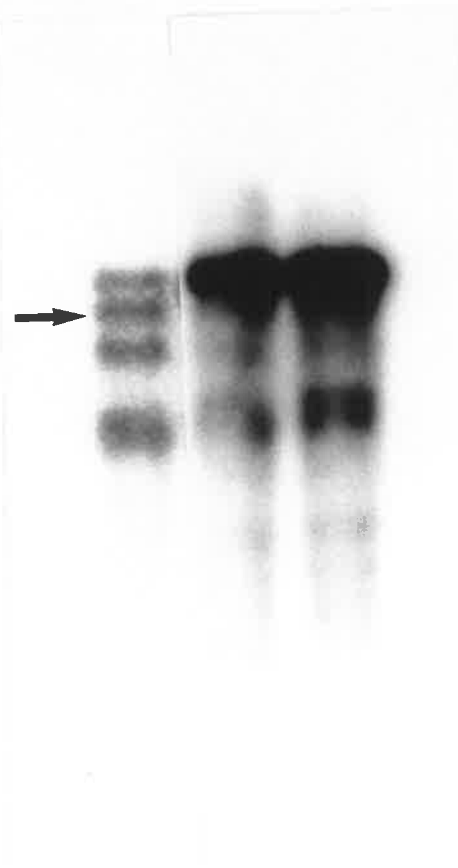
1) a polyadenylated RNA enriched fraction.
2) 1 µg of a polyadenylated RNA depleted fraction.
3) 1 µg of beet total RNA.

B) Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to RNA isolated from tobacco green leaf tissue. RNA samples consisted of:

1) a polyadenylated RNA enriched fraction.
2) 1 µg of a polyadenylated RNA depleted fraction.
3) 1 µg of tobacco total RNA.

An arrow indicates a RNA transcript that shows homology to the plastid DNA probe and can only be identified in the polyadenylated RNA enriched fraction.

C) Lighter exposure of the Northern blot membrane depicted in B).

A
1 2 3

B
1 2 3

C
1 2 3

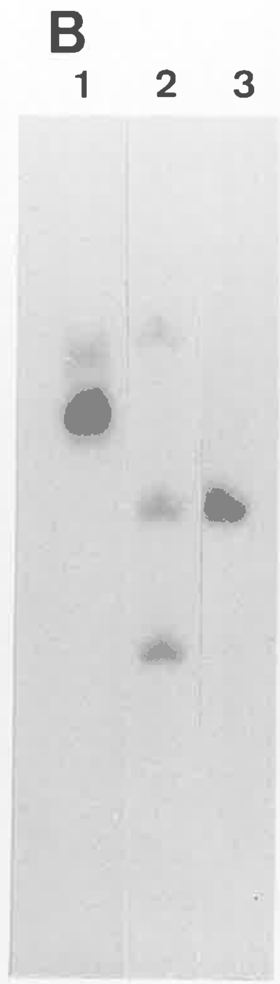
Figure 7.2: A lambda gt10 cDNA clone with plastid DNA homology.

A) Agarose gel showing restriction products of DNA isolated from a lambda gt10 cDNA clone with homology to the spinach plastid 7.7 kbp PstI DNA fragment, upon restriction with 1) EcoRI, 2) BamHI and 3) HindIII. Track 4 contained 0.5 µg of wild type lambda DNA restricted with HindIII.

B) Hybridization of the spinach plastid 7.7 kbp PstI DNA fragment to a Southern blot of the agarose gel depicted in A).

C) Agarose gel showing restriction products of DNA isolated from four single plaques of the lambda gt10 cDNA clone with plastid DNA homology. Tracks 1-3, 4-6, 7-9 and 10-12 contained DNAs isolated from single plaques. DNAs were restricted with EcoRI (tracks 1, 4, 7 and 10), BamHI (tracks 2, 5, 8 and 11) and HindIII (tracks 3, 6, 9 and 12).

A
1 2 3 4


B
1 2 3

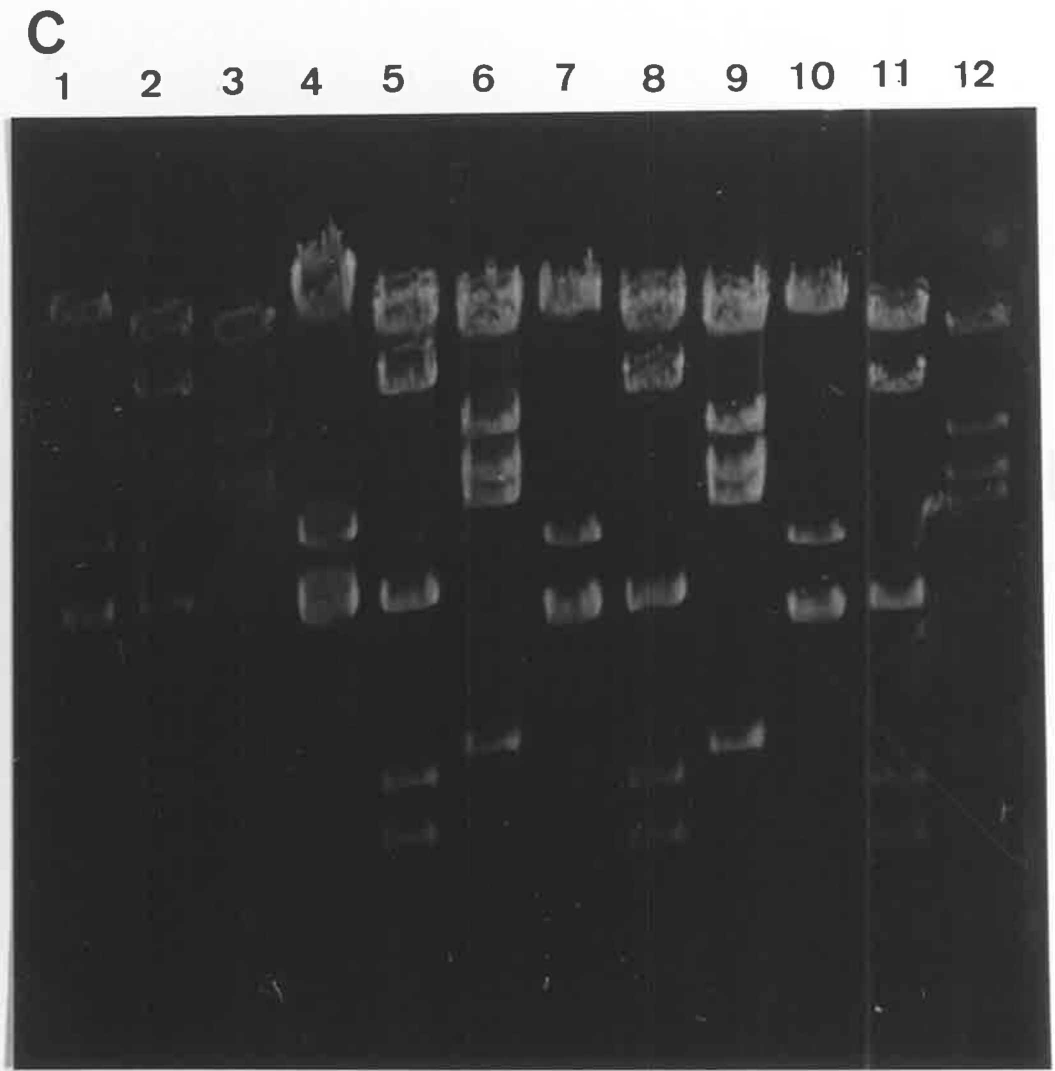

C
1 2 3 4 5 6 7 8 9 10 11 12

Figure 7.3: Gene map of the 4.4 kbp HindIII fragment of the lambda gt10 cDNA clone with plastid DNA homology.

The 4.4 kbp HindIII fragment of the lambda cDNA clone is represented by a light green line. The plastid genes encoded by this fragment are boxed in dark green and appropriately labelled. Genes labelled s, g and fM correspond to plastid genes *trn*S, *trn*G and *trn*fM. The position of HindIII and EcoRI restriction sites on this fragment are labelled "H" and "E" respectively. Arrows indicate regions of this fragment that have been sequenced. Numbers correspond to the numbering of the published tobacco plastid genome sequence (Shinozaki *et al*, 1986).

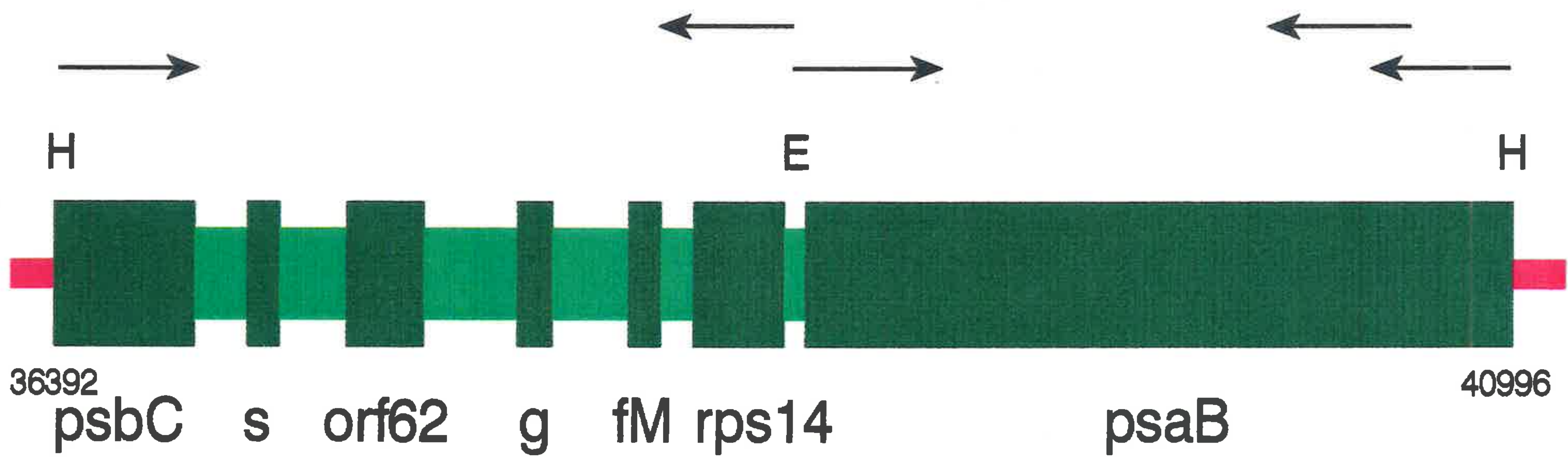
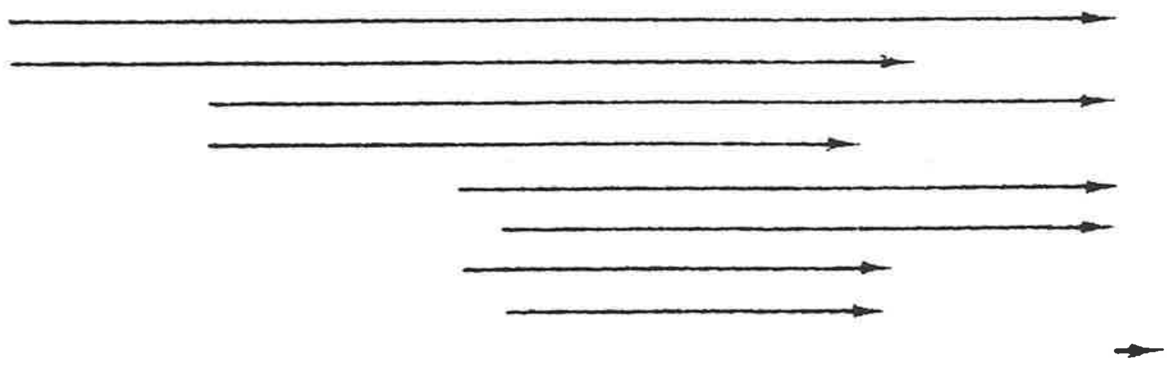
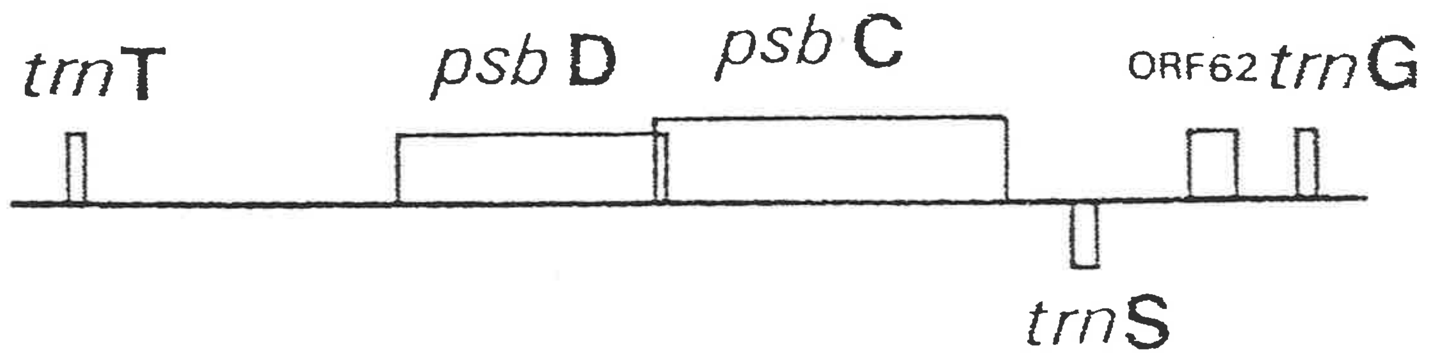

H
E
H
36392
40996
psbC
s
orf62
g
fM
rps14
psaB

Figure 7.4: Transcription of the tobacco plastid *psb*D-*psb*C gene cluster.

Extracted directly from Yao *et al*, 1989.

Detection of transcripts from *trn*T, *psb*D, *psb*C, *trn*S, ORF62 and *trn*G. Arrows indicate the extent of proposed transcripts of this gene cluster and the size of each transcript is given in kbp. Genes are boxed and appropriately labelled.


4.4
3.6 a
3.6 b
2.6 a
2.6 b
2.4
1.7
1.5
0.3


trnT
psb D
psb C
ORF62
trnG
trnS

CHAPTER 8

Discussion

This study has further demonstrated that plastid DNA sequence homologies in the nuclear genomes of higher plants is a widespread phenomenon involving both dicotyledenous and monocotyledenous plant species. In some species these nuclear sequences appear highly conserved while other species (eg. *Beta vulgaris*) show sequence heterogeneity possibly ranging to somatic instability. Some of this variation may be attributable to different amounts of heterozygo sity being present in populations of these plant species, resulting from differences in the breeding systems of each species and the diversity of the genetic base from which they were derived. These promiscuous plastid sequences may be useful as a rather general indicator of genetic diversity within a plant population or species. Most agricultural plant species are produced by selective inbreeding and it is important to assess the germplasm diversity in these species to maintain both yield and pathogen immunity.

The apparent somatic variability of plastid sequences within the nuclear genome of *B. vulgaris* represents the upper extreme of promiscuous sequence heterogeneity observed. The integration of these foreign sequences into the plant nuclear genome may induce localized instability of these regions as has been reported for other species (Murnane, 1990). Promiscuous sequences present within the nuclear genomes of rice (Kikuchi *et al*, 1987) and *Cucumis melo* (Grisvard *et al*, 1990) have been shown to be unstable during cellular differentiation. The tissue specific variation of nuclear sequences with plastid DNA homology in *B. vulgaris*, could also be the result of ongoing transposition of plastid sequences to the nucleus.

The different amounts of plastid sequence divergence of clones 1M and 3D demonstrate, that at least in evolutionary terms, transposition of plastid sequences to the tobacco nuclear genome may be an ongoing process. The 3D and 1M sequences were estimated to represent transposition events that took place 700,000 and 8 million years ago respectively. This result parallels the identification of mitochondrial DNA inserts in

the human nuclear genome which have estimated transposition ages ranging from less than 9 to 40 million years ago (Fukuda *et al*, 1985; Nomiyama *et al*, 1984; 1985; Kamimura *et al*, 1989).

No plastid sequence homologies could be detected within the nuclear genome of *Arabidopsis thaliana*, suggesting that these plastid sequences are part of nonessential, probably repetitive regions of the genome. The mapping of plastid sequence homologies to chromosomes 5, 9 and 10 in tomato (Pichersky *et al*, 1991; Pichersky and Tanksley, 1988) suggest that these homologies are distributed throughout the plant nuclear genome in some species. Adjacent to the plastid sequence in clone 1M was a member of a family of dispersed repeats that were extensively methylated in the tobacco nuclear genome.

Characterised in this study are the largest promiscuous sequences identified to date. The tobacco nucleus was shown to contain large tracts of plastid DNA homology in excess of 18 kbp. At least one third of the plastid genome is represented in the tobacco nucleus in these large contiguous tracts of DNA. Determination of the absolute size of these plastid integrants awaits further investigation by either pulse field gel electrophoresis or YAC cloning. Promiscuous plastid sequences have been identified in maize YAC libraries but their absolute size has not been determined (Gupta and Hoo, 1991). Limited empirical calculations based upon a random distribution of Sau3A cloning sites in nuclear DNA suggests that finding only one nuclear/plastid DNA junction in a total of almost 100 kbp of cloned DNA indicates the presence of nuclear integrants of plastid DNA in excess of 155 kbp. This suggests the possibility of entire plastid genomes in the nuclear DNA of tobacco.

The large size of plastid DNA integrants in the nuclear genome of tobacco contrasts other results where sequence analyses are available (Cheung and Steele Scott, 1989; Pichersky *et al*, 1991; Pichersky and Tanksley, 1988). The most conspicuous difference from these observations occurs in tomato where, in a systematic search for plastid sequences integrated into the nuclear genome, only two very short integrants were isolated. However the cloning strategy employed by Pichersky *et al* (1991) would

almost have certainly excluded the large contiguous tracts of plastid DNA identified in this study. As this search of two clone libraries covered 95% of the tomato nuclear genome and 33% of the plastid genome, both the rarity and the small size of integrants in tomato contrast the arrangement of such sequences in tobacco. The nuclear genome of tobacco ($4.8x10^6$ kb, Bennett and Smith, 1976) is much larger than that of tomato ($7x10^5$ kb, Pichersky and Tanksley, 1988) and the former is a tetraploid whereas the latter is a diploid; again suggesting that genome size may influence the presence and abundance of these sequences within a plant species. Both species are members of the Solonaceae and are therefore closely related but there are differences in the way these two genomes are organized which is reflected in these contrasting results. Hybridization with poly(dA) (Pichersky *et al*, 1990) gives a very strong signal to Southern blots of tobacco DNA and almost none to tomato, an observation which was interpreted to indicate the presence of many more processed pseudogenes in tobacco.

An alternative explanation for the isolation of these large tracts of plastid DNA (represented by clones 3D, 3-2, 3b, 1H and 2d) is that a subpopulation of heavily methylated plastid genomes is present in tobacco. Plastid DNA methylation has been identified in some plants, but this methylation is very limited involving only a few specific sites (Ngernprasirtsiri *et al*, 1988a; 1988b; 1989). Furthermore this plastid population would have to copurify with plant nuclei and be genomically highly polymorphic to account for both the identification of these sequences in a tobacco nuclear DNA preparation in addition to the sequence differences observed between clones 3D, 3-2 and 3b. All plastids are generally derived from a small number of maternal proplastids in the fertilized egg, so it is difficult to envisage the formation of this amount of sequence variation within a plastid population. Of the three clones that were most extensively sequenced in this study, all encoded essential plastid genes that were either inactivated or altered as a result of sequence divergence when compared to the true plastid genome. This evidence therefore strongly suggests that these large tracts of plastid DNA are of nuclear origin and not part of any functional plastid genome.

In addition to being large these tobacco nuclear sequences are also present in a relatively high copy number. The tobacco nuclear genome contains multiple copies of these plastid integrants, with the number of different integrants ranging from 15 to 57 copies per 1c tobacco nucleus. These copy number estimates are significantly larger than those obtained for spinach, where an estimated five copies of the plastome are present per haploid nuclear genome (Steele Scott and Timmis, 1984), and the tomato nuclear genome where apparently very few promiscuous plastid sequences are present (Pichersky *et al*, 1991). Examination of plastid sequence homologies in the spinach (Cheung and Steele Scott, 1989; unpublished results) and tomato (Pichersky and Tanksley, 1988; Pichersky *et al*, 1991) nuclear genome has indentified only small plastid DNA inserts. Therefore a relationship between size and copy number of plastid DNA integrants may exist. The much larger promiscuous plastid DNA copy number obtained for tobacco parallels estimates of over several hundred mitochondrial DNA integrants within the human (Fukuda *et al*, 1985; Kamimura *et al* 1989) and locust (Gellisson, 1983; Gellisson and Michaelis, 1988) nuclear genomes.

The multicopy status of these nuclear sequences may not directly correlate with the number of plastid transposition events that took place. The identification of 15 copies of plastid DNA in the tobacco nucleus, each containing the same 41 bp deletion present in clone 3D, suggests that they all arose from a duplication of a single transposition event. Analysis of the sites of integration of most of these sequences may therefore not yield information on the mechanism of sequence transfer of these plastid sequences to the nucleus, but rather represent secondary nuclear integration events. A variety of other promiscuous integrants also show evidence of sequence duplication or rearrangement (Farrelly and Butow, 1983; Wright and Cummings, 1983; Pichersky and Tanksley, 1988; Jacobs *et al*, 1983, 1986; Nomiyama *et al*, 1984; Wakasugi *et al*, 1985; Kikuchi *et al*, 1987; Grisvard *et al*, 1990). Some of these rearrangements are so extreme that a chimeric promiscuous sequence can be produced consisting of combinations of organelle sequences that are widely separated upon the organelle genome (Moon *et al*, 1988; Cheung and Steele Scott, 1989; Ossorio *et al*, 1991; Kamimura *et al*, 1989). The

absence of any unifying sequence motifs at the integration sites of all promiscuous nuclear sequences identified to date may be the result of post-transposition nuclear duplications and rearrangements. Alternatively nuclear integration of these sequences may be a random event analogous to SV40 and Ti plasmid integration into nuclear genomes (Gheysen *et al*, 1991; 1987; Roth and Wilson, 1986).

Analysis of plastid/nuclear DNA junctions is further complicated by not knowing the initial sequence of the nuclear site of integration. Integration of foreign sequences into nuclear and mitochondrial genomes can involve both target site deletion and the addition of many nucleotides (so called filler DNA) at the site of integration (Roth *et al*, 1989; Gheysen *et al*, 1987; Sainsard-Chanet and Begel, 1990; Landau *et al*, 1987). Recent advances in plastid transformation may soon make it possible to overcome these problems (Blowers *et al* 1990; Haring and DeBlock, 1990; Newman *et al*, 1990; Sporlein *et al*, 1991). Engineered genes could be inserted into the plastid genome that encoded selectable markers expressed only in a nuclear environment. A transposition event from the plastid genome to the nuclear genome could therefore be selected for. The sequence of the nuclear target site both prior and post-transposition could thus be determined.

The sequence divergence between these plastid homologies and the *bona fide* plastid genome may give insights into nuclear mutation processes. Several of the lambda clones in this study had deletions and insertions associated with repeat motifs suggesting that processes such as DNA replication slippage and excision of hairpin loop structures may have been involved in their formation. The high sequence conservation of the plastid genome implies that most of the differences between a promiscuous plastid sequence and the *bona fide* plastid genome will be the result of nuclear mutation events.

The large tracts of nuclear DNA with homology to the plastid genome cloned in this study did not contain flanking nuclear regions so little information can be derived concerning possible mechanisms of transfer and integration of these five sequences. However as these integrants are very large and contain nontranscribed regions of the

plastid genome their transposition to the nucleus would not appear to have been mediated by an RNA intermediate, as has been suggested for some promiscuous sequences (Schuster and Brennicke, 1987; Nugent and Palmer, 1991). The identification of chimeric promiscuous sequences, that consist of combinations of organelle sequences that are widely separated upon the same organelle genome (Moon *et al*, 1988; Cheung and Steele Scott, 1989; Ossorio *et al*, 1991; Kamimura *et al*, 1989) also suggests that a single RNA intermediate was not involved in transposition of these sequences. Conversely sequence analysis of a plant nuclear encoded *cox*II gene demonstrates that organelle sequences can be transferred to the nuclear genome *via* an RNA intermediate (Nugent and Palmer, 1991). These contrasting results support the hypothesis that at least two distinct molecular mechanisms are involved in the transfer of organelle sequences to the plant nuclear genome.

REFERENCES

Albertini A.M., Hofer M., Calos M.P., Miller J.H. (1982). On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell* 29: 319-328.

Allard R.W. (1960). Principles of plant breeding. John Wiley and Sons Inc., New York.

Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seidmann J.G., Smith J.A., Struhl K. (1987). Current protocols in molecular biology. John Wiley and Sons Inc., New York.

Ayliffe M.A., Timmis J.N., Steele Scott N. (1988). Homologies to chloroplast DNA in the nuclear DNA of a number of Chenopod species. *Theor. Appl. Genet.* 75: 282-285.

Baldauf S.L., Manhart J.R., Palmer J.D. (1990). Different fates of the chloroplast tufA gene following its transfer to the nucleus in green algae. *Proc. Natl. Acad. Sci. USA* 87(14): 5317-5321.

Bauldauf S.L., Palmer J.D. (1990). Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. *Nature* 344: 262-265.

Bayer M.G., Schenk H.E.A. (1989). Ferredoxin of *Cyanophora paradoxa* Korsch is encoded on cyanellar DNA . *Curr. Genet.* 16(4): 311-313.

Belfort M. (1991). Self splicing introns in procaryotes: migrant fossils ? *Cell* 64: 9-11.

Bendich A.J., Smith S.B. (1990). Moving pictures and pulse-field electrophoresis show linear DNA molecules from chloroplast and mitochondria. *Curr. Genet.* 17: 421-425.

Bennet M.D., Smith J.D. (1976). Nuclear DNA amounts in angiosperms. *Philo. Trans. Roy. Soc. (Lon.) B* 274: 227-274.

Bernatzky R., Mau S-L., Clarke A.E. (1989). A nuclear sequence associated with self incompatibility in *Nicotiana alata* has homology with mitochondrial DNA. *Theor. Appl. Genet.* 77: 320-324.

Bertrand H., Chan B.S., Griffiths A.J.F. (1985). Insertion of a foreign nucleotide sequence into mitochondrial DNA causes senescence in *Neurospora intermedia. Cell* 41: 877-884.

Bettini P., MacNally S., Sevignac M., Dron M. (1988). A mitochondrial transcript with homology to the 3' end of the chloroplast *psb*A gene is present only in the atrizine resistant biotype of *Chenopodium album. Theor. Appl. Genet.* 75: 291-297.

Bloom G.D. (1967). A nucleus with cytoplasmic features. *J. Cell Biol.* 35: 266-268.

Blowers A.D., Ellmore G.J., Klein U., Bogorad L. (1990). Transcriptional analysis of endogenous and foreign genes in chloroplast transformants of *Chlamydomonas. The Plant Cell* 2: 1059-1070.

Blumenthal R.M. (1989). Cloning and restriction of methylated DNA in *Escherichia coli. Focus* 11(3): 41-46.

Boer P.H., Bonen L., Lee R.W., Gray M.W. (1985). Genes for respiratory chain proteins and ribosomal RNA's are present on a 16 kilobase-pair DNA species from *Chlamydomonas reinhardtii* mitochondria. *Proc. Natl. Acad. Sci. USA* 82: 3340-3344.

Brandes D., Schofield B.H., Anton E. (1965). Nuclear mitochondria ? *Science* 149: 1373-1374.

Braun C.J., Levings III C.S. (1986). DNA sequence homology between the maize mitochondrial and chloroplast genomes. *Maize Genet. Coop. News.* 60: 111.

Breiman A., Galun E. (1990). Nuclear-mitochondrial interrelations in angiosperms. *Plant Sci.* 71: 3-19.

Brinkmann H., Martinez P., Quigley F., Martin W., Cerff R. (1987). Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Evol.* 26: 320-328.

Brown A.H.D. (1978). Isozymes, plant population genetic structure and genetic conservation. *Theor. Appl. Genet.* 52: 145-157.

Brown T.A., Ray J.A., Waring R.B., Scazzocchio C., Davies R.W. (1984). A mitochondrial reading frame which may code for a second form of ATPase subunit 9 in *Aspergillus nidulans. Curr. Genet.* 8: 489-492.

Burgess D.G., Taylor W.C. (1988). The choroplast affects the transcription of a nuclear gene family. *Mol. Gen. Genet.* 214 :89-96.

Carlson J.E., Erickson L.R., Kemble R.J. (1986). Cross hybridization between organelle RNA's and mitochondrial and chloroplast genomes in *Brassica. Curr. Genet.* 11: 161-163.

Cattaneo R. (1992). RNA editing: in chloroplast and brain. *Trends Bio. Sci* 17: 4-5.

Cheung W.Y., Steele Scott N. (1989). A contiguous sequence in spinach nuclear DNA is homologous to three separated sequences in chloroplast DNA. *Theor. Appl. Genet.* 77: 625-633.

Cornelissen M., Vandewiele M. (1989). Nuclear transcriptional activity of the tobacco plastid *psb*A promoter. *Nucl. Acids Res.* 17(1): 19-29.

Corral M., Baffet G., Kitzis A., Paris B., Tichonicky L., Kruh J., Guguen-Guillouzo C., Defer N. (1989). DNA sequences homologous to mitochondrial genes in nuclei from normal rat tissue and from rat hepatoma cells. *Biochem. Biophys. Res. Commun.* 162: 258-264.

Covello P.S., Gray M.W. (1989). RNA editing in plant mitochondria. *Nature* 341: 662-666.

Crouzillat D., Gentzbittel L., de la Canal L., Vaury C., Perrault A., Nicolas P., Ledoigt G. (1989). Properties and nucleotide sequence of a mitochondrial plasmid from sunflower. *Curr. Genet.* 15: 283-289.

Curtis S.E., Rawson J.R.Y. (1982). Nucleotide sequence homology exists between the chloroplast and nuclear ribosomal DNA's of *Euglena gracilis. Plant Physiol.* 69: 67-71.

de Goer S.L., Markowicz Y., Dalmon J., Andren H. (1988). Physical maps of two circular plastid DNA molecules of the brown alga *Pylaiella littoralis* (L) Kjellm. *Curr. Genet.* 14: 155-162.

Deikmann C.L., Gandy B. (1987). Preferential recombination between GC clusters in yeast mitochondrial DNA. *EMBO* 6: 4197-4203.

Delihas N., Fox G.E. (1988). Origins of the plant chloroplasts and mitochondria based on comparisons of 5s ribosomal RNA's. *Ann. N. Y. Acad Sci.* 503: 92-101.

Deng X-W., Wing R.A., Gruissem W. (1989). The chloroplast genome exists in multimeric forms. *Proc. Natl. Acad. Sci. USA* 86: 4156-4160.

Dennis E.S., Finnegan E.J., Taylor B.H., Peterson T.A., Walker A.R., Peacock W.J. (1988). Maize transposable elements: Structure, function and regulation. In: Plant transposable elements. Nelson O. (Ed.). Plenum Press, New York.

De Pamphilis C.W., Palmer J.D. (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348: 337-339.

Devereux J., Haeberli P., Smithies O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12: 387-395.

Dewey R.E., Levings C.S., Timothy D.H. (1986). Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male sterile cytoplasm. *Cell* 44: 439-449.

Doerfler W., Toth M., Kochanek S., Achten S., Freisem-Rabien U., Behb-Krappa A., Orend G. (1990). Eucaryotic DNA methylation. Facts and problems. *FEBS Lett.* 268: 329-333.

Douglas S.E., Murphy C.A., Spencer D.F., Gray M.W. (1991). *Cryptomonad* algae, evolutionary chimeras of two phylogenetically distinct unicellular eucaryotes. *Nature* 350: 148-151.

Drake J.W., Glickman B.W., Ripley L.S. (1983). Updating the theory of mutation. *Am. Sci.* 71: 621-630.

Dron M., Hartmann C., Rode A., Sevignac M. (1985). Gene conversion as a mechanism for divergence of a chloroplast tRNA gene inserted in the mitochondrial genome of *Brassica oleracea*. *Nucl. Acids Res.* 13: 8603-8610.

du Jardin P. (1990). Homologies to plastid DNA in the nuclear and mitochondrial genomes of potato. *Theor. Appl. Genet.* 79: 807-812.

Ellis J. (1982). Promiscuous DNA-chloroplast genes inside plant mitochondria. *Nature* 299: 678-679.

Ellis J. (1987). Proteins as molecular chaperons. *Nature* 328: 378-379.

Fangman W.L., Henly J.W., Churchill G., Brewer B.J. (1989). Stable maintenance of a 35 base-pair yeast mitochondrial genome. *Mol. Cell. Biol.* 9: 1917-1921.

Farrelly F., Butow R.A. (1983). Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* 301: 296-301.

Fearnely I.M., Walker J.E. (1986). Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. *EMBO* 5: 2003-2008.

Fejes E., Masters B.S., McCarthy D.M., Hauswirth W.W. (1988). Sequence and transcriptional analysis of a chloroplast insert in the mitochondrial genome of *Zea mays*. *Curr. Genet.* 13: 509-515.

Flavell R.B. (1988). DNA transposition - a major contributor to plant chromosome structure. *Bioessays* 1: 21-22.

Flugge U-I. (1990). Import of proteins into chloroplasts. *J. Cell Sci.* 96: 351-354.

Frischauff A.M., Lehrach H., Poustka A., Murray N. (1983). Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170: 827-842.

Fukuda M., Wakasugi S., Tsuzuki T., Nomiyama H., Shimada K. (1985). Mitochondrial DNA-like sequences in the human nuclear genome: Characterisation and implications in the evolution of mitochondrial DNA. *J. Mol. Biol.* 186: 257-266.

Gantt J.S., Baldauf S.L., Calie P.J., Weeden N.F., Palmer J.D. (1991). Transfer of *rpl*22 to the nucleus greatly proceeded its loss from the chloroplast and involved the gain of an intron. *EMBO* 10: 3073-3078.

Gauly A., Kossel H. (1989). Evidence for tissue-specific cytosine-methylation of plastid DNA from *Zea mays*. *Curr. Genet.* 15: 371-376.

Gellissen G., Bradfield J.Y., White B.N., Wyatt G.R. (1983). Mitochondrial DNA sequences in the nuclear genome of a locust. *Nature* 301: 631-634.

Gellissen G., Michaelis G. (1988). Gene transfer: mitochondria to nucleus. *Ann. N.Y.Acad. Sci.* 503: 391-401.

Gheysen G., Montagu M.V., Zambryski P. (1987). Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangement of target plant DNA sequences. *Proc. Natl. Acad. Sci. USA*: 6169-6173.

Gheysen G., Villarrod R., Montagu M.V. (1991). Illegitimate recombination in plants: a model for T-DNA integration. *Genes and Developmen*t 5: 287-297.

Goodenough U.W. (1989). Basal body chromosomes. *Cell* 59: 1-3.

Gray M.W. (1989a). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* 5: 25-50.

Gray M.W. (1989b). The evolutionary origin of organelles. *TIG* 5: 294-299.

Grenbaum Y., Naveh-Many T., Cedar H., Razin A. (1981). Sequence specificity of methylation in higher plant DNA. *Nature* 292: 21-26.

Grisvard J., Sevignac M., Chateau M., Branchard M. (1990). Changes in a repetitive DNA sequence during callus culture of *Cucumis melo*. *Plant Science* 72: 81-91.

Gruissem W. (1989). Chloroplast gene expression: How plants turn their plastids on. *Cell* 56: 161-170.

Gualberto J.M., Lamattina L., Bonnard G., Weil J-H., Greinenberger J-M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341: 660-662.

Gubler U., Hoffman B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263-269.

Gupta M., Hoo B. (1991). Entire maize chloroplast chromosome is stably maintained in a yeast artificial chromosome. *Plant Mol. Biol.* 17: 361-369.

Hadler H.I., Dimitrijevic B., Mahalingam R. (1983). Mitochondrial DNA and nuclear DNA from normal rat liver have a common sequence. *Proc. Natl. Acad. Sci. USA* 80: 6495-6499.

Hall J.L., Ramanis Z., Luck D.J.L. (1989). Basal body/centriolar DNA: Molecular genetic studies in *Chlamydomonas*. *Cell* 59: 121-132.

Hallick R.B., Bottomley W. (1983). Proposals for the naming of chloroplast genes. *Plant Mol. Biol. Rep.* 1: 38-43.

Hancock K., Hajduk S.L. (1990). The mitochondrial tRNA's of *Trypanosoma brucei* are nuclear encoded. *J. Biol. Chem.* 265: 19208-19215.

Hanley-Bowdoin L., Orozco E.M., Chua N-H. (1985). Transcription of chloroplast genes by homologous and heterologous RNA polymerases. In Molecular Biology of the Photosynthetic Apparatus. Cold Spring Harbour.

Haring M.A., DeBlock M. (1990). New roads towards chloroplast transformation in higher plants. *Physiol. Plant.* 79: 218-220.

Heinhurst S., Shively J.M. (1983) Encoding of both subunits of ribulose-1,5-bisphosphate carboxylase by the organelle genome of *Cyanophora paradoxa*. *Nature* 304: 373-374.

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 chaperon oligomeric protein assembly. *Nature* 333: 330-334.

Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M., Mori M., Kondo C., Honji Y., Sun C-R., Meng B-Y., Li Y-Q., 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.

Hoch B., Maier R.M., Appel K., Igloi G.L., Kossel H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353: 178-180.

Hudson A.P., Lilijestrand-Golden C., Butow R.A. (1985). Mitochondrial DNA sequences in the yeast nuclear genome.
U.C.L.A. Symp. Mol. Cell. Biol. 20: 15-26. (Ed.) Liss A.R.

Hughes E. (1990). Approaches to developing probes for DNA fingerprint analysis in citrus. Honours thesis. University of Adelaide, Genetics Department.

Huynh T.V., Young R.A., Davis R.W. (1985). Constructing and screening cDNA clone libraries in *lambda* gt10 and *lambda* gt11. In: DNA cloning techniques: A practical approach. Glover D. M. (Ed.) I.R.L. Press, Oxford, U.K.

Hvid S., Nielson G. (1977). Esterase isozyme variants in barley. *Hereditas* 87: 155-162.

Hwang S.R., Tabita R.F. (1991). Acyl carrier protein-derived sequence encoded by the chloroplast genome in the marine diatom *Cylindrotheca sp.* strain N1. *J. Biol Chem.* 266: 13492-13494.

Iams K.P., Heckman J.E., Sinclair J.H. (1985). Sequence of histidyl tRNA present as a chloroplast DNA insert in mitochondrial DNA of *Zea mays*. *Plant Mol. Biol.* 4: 225-232.

Infante D. H., Weissbach A. (1990). Organellar DNA replication in *Nicotiana tabacum* cultured cells. *Plant Mol. Biol.* 14: 891-897.

Jacobs H.T., Grimer B. (1986). Complete nucleotide sequences of the nuclear psuedogenes for cytochrome oxidase subunit I and the large mitochondrial rRNA in the sea urchin *Strongylocentrotus purpuratus. J. Mol. Biol.* 187: 509-527.

Jacobs H.T., Posakony J.W., Grula J.W., Roberts J.W., Xing J.H., Britten R.J., Davidson E.H. (1983). Mitochondrial DNA sequences in the nuclear genome of *Strongylocentrotus purpuratus. J. Mol. Biol.* 165: 609-632.

Jamet-Vierny C., Begel O., Belcour L. (1980). Senescence in *Podospora anserina*: Amplification of a mitochondrial DNA sequence. *Cell* 2: 189-194.

John B., Miklos G. (1988). The eucaryote genome in development and evolution. Unwin Hymand Ltd, London.

Jolley E., Smith D.C. (1980). The green hydra symbiosis II: The biology of establishment of the association. *Proc. R. Soc. (Lon.) B* 207: 311-333.

Joyce P.B.M., Gray M.W. (1989). Chloroplast-like transfer RNA genes expressed in wheat mitochondria. *Nucl. Acids Res.* 17: 5461-5476.

Jubier M.F., Lucas H., Delcher E., Hartmann C., Quetier F, Lejeune B. (1990). An internal part of the chloroplast *atp*A gene sequence is present in the mitochondrial genome of *Triticum aestivum*: molecular organization and evolutionary aspects. *Curr. Genet.* 17: 523-528.

Kamimura N., Ishii S., Liandong M., Shay J.W. (1989). Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. *J. Mol. Biol.* 210: 703-707.

Kanazawa A., Sakamoto W., Kishimoto N., Yano M., Tsutsumi N., Saito A., Tano S. (1991). Linkage analysis of the nuclear homologues of mitochondrial plasmid-like DNA's in rice. *Jnp. J. Genet.* 66: 597-607.

Kaiser K., Murray N. (1985). The use of phage lambda replacement vectors in construction of representative genomic libraries. In: DNA cloning. A practical approach (vol I.). D. M. Glover (ed.).IRL press.

Kavanagh T.A., Timmis J.N. (1986). Heterogeneity in cucumber ribosomal DNA. *Theor. Appl. Genet.* 72: 337-345.

Kemble R.J., Bedbrook J.R. (1980). Low molecular weight circular and linear DNA in mitochondria from normal and male-sterile *Zea mays* cytoplasm. *Nature* 284: 565-566.

Kemble R.J., Mans R.J., Gabay-Laughnan S., Laughnan J.R. (1983). Sequences homologous to episomal mitochondrial DNA's in the maize nuclear genome. *Nature* 304: 744-747.

Kim W.K., Klassen G.R. (1989). Double stranded RNA's in mitochondrial extracts of stem rusts and leaf rusts of cereals. *Curr. Genet.* 15: 161-166.

Kikuchi S., Takaiwa F., Oono K. (1987). Variable copy number DNA sequences in rice. *Mol. Gen. Genet.* 210: 373-380.

Koll F. (1986). Does nuclear integration of mitochondrial sequences occur during senescence of *Podospora. Nature* 324: 597-599.

Kuhsel M.G., Strickland R., Palmer J.D. (1990). An ancient group I intron shared by eubacteria and chloroplasts. *Science* 250: 1570-1573.

Kuiper M.T.R., Lambowitz A.M. (1988). A novel reverse transcriptase activity associated with mitochondrial plasmids of *Neurospora. Cell* 55: 693-704.

Kung S.D., Lin C.M. (1985). Chloroplast promoters from higher plants. *Nucl. Acids Res.* 13: 7543-7549.

Lacoste-Royal G., Gibbs P. (1985). *Ochromonas* mitochondria contains a specific chloroplast protein. *Proc. Natl. Acad. Sci. USA* 82: 1456-1459.

Lambowitz A.M. (1989). Infectious introns. *Cell* 56: 323-326.

Landau N.R., Schatz D.G., Rosa M., Baltimore D. (1987). Increased frequency of N-region insertion in a murine pre-B-cell line infected with terminal deoxynucleotidyl transferase retroviral expression vector. *Mol. Cell. Biol.* 7: 3237-3243.

Lestienne P. (1989). Mitochondrial and nuclear DNA complementation in the respiratory chain function and defects. *Biochimie* 71: 1115-1123.

Levings III C.S., Brown G.G. (1989). Molecular biology of plant mitochondria. *Cell* 56: 171-179.

Li W-H., Luo C-C., Wu C-I. (1985). Evolution of DNA sequences. In: Molecular Evolutionary Genetics. MacIntyre R.J.(Ed.). Plenum, New York.

Li-Weber M., Leible M., Schweiger M. (1989). Difference in location in Dasycladaceae of a DNA sequence homologous to the *Drosphilia* per locus. *Plant Cell Rep.* 8: 169-173.

Liaud M-F., Zhang D.X., Cerff R. (1990). Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus. *Proc. Natl. Acad. Sci. USA* 87: 8918-9822.

Lonsdale D.M. (1984). A review of the structure and organization of the mitochondrial genome in higher plants. *Plant Mol. Biol.* 3: 201-206.

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.

Lundquist A., Osterbye U., Larsen K., Linde-Laursen I.B. (1973). Complex self-incompatability systems in *Ranunculus acris* L. and *Beta vulgaris* L. *Hereditas* 74: 161-168.

Makaroff C.A., Palmer J.D. (1987). Extensive mitochondrial specific transcription of the *Brassica campestris* mitochondrial genome. *Nucl. Acids Res.* 15: 5141-5156.

Maniatis T., Fritsch E.F., Sambrook J. (1982). Molecular cloning. A laboratory manual. Cold Spring Harbour Laboratory Publications, U.S.A.

Marano M., Carrillo N. (1991). Chromoplast formation during tomato fruit ripening. No evidence for plastid DNA methylation. *Plant Mol. Biol.* 16: 11-19.

Marechal L., Runeberg-Ross P., Grienenberger J.M., Colin J., Weil J.H., Lejeune B., Quetier F., Lonsdale D.M. (1987). Homology in the region containing a $tRNA^{trp}$ gene

ADELAIDE UNIVERSITY LIBRARY

and a (complete or partial) $tRNA^{pro}$ gene in wheat mitochondrial and plastid genomes. *Curr. Genet.* 12: 91-98.

Marechal-Drouard L., Guillemaut P., Cosset A., Arbogast M., Weber F., Weil J.H., Dietrich A. (1990). Transfer RNA's of potato (*Solanum tuberosum*) mitochondria have different genetic origins. *Nucl. Acids Res.* 18: 3689-3696.

Margulis L. (1970). Origin of eucaryotic cells . Yale University Press. New Haven, Conn.

Markowitz Y., deGoer S.L., Mache R. (1988). Presence of a 16s rRNA pseudogene in the bi-molecular plastid genome of the primitive brown algae *Pylaiella littoralis*. Evolutionary implications. *Curr. Genet.* 14:599-608.

Marshall D.R., Allard R.W. (1970). Maintenance of isozyme polymorphisms in natural populations of *Avena barbata. Genetics* 66: 393-399.

Martin W., Cerff R. (1986). Procaryotic features of a nuclear-encoded enzyme. *Eur. J. Biochem.* 159: 323-331.

Martin W., Lagrange T., Li Y.F., Bisanz-Seyer C., Mache R. (1990). Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. *Curr. Genet.* 18: 553-556.

Meinhardt F., Kempken F., Kamper J., Esser K. (1990). Linear plasmids among eucaryotes: fundamentals and application. *Curr. Genet.* 17: 89-95.

Messing J. (1979). A multipurpose cloning system based on single-stranded DNA bacteriophage M13. *Rec. DNA Tec. Bull.* 2: 43.

Meyerowitz E.M., Pruitt R.E. (1985). *Arabidopsis thaliana* and plant molecular genetics. *Science* 229: 1214-1218.

Milligan B.G., Hampton J.N., Palmer J.D. (1989). Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* 6: 355-368.

Moon E., Kae T-H., Wu R. (1988). Rice mitochondrial genome contains a rearranged chloroplast gene cluster. *Mol. Gen. Genet.* 213: 247-253.

Morden C.W., Golden S.S. (1989). *psb*A genes indicate common ancestory of prochlorophytes and chloroplasts. *Nature* 337: 382-384.

Murnane J.P. (1990). The role of recombinational hotspots in genome instability in mammalian cells. *Bioessays* 12: 557-581.

Murnane J.P., Young B.R. (1989). Nucleotide sequence analysis of novel junctions near an unstable integrated plasmid in human cells. *Gene* 84: 201-205.

Nagley P. (1989). Trafficking in small mitochondrial RNA molecules. *TIG* 5: 67-69.

Nagley P. (1991). Coordination of gene expression in the formation of mammalian mitochondria. *TIG* 7: 1-4.

Nags D.K., Petes T.D. (1991). Even-base-pair inverted repeats in DNA form stable hairpins *in vivo* in *Saccharomyces cerevisiae*. *Genet.* 129: 669-673.

Neale A.D., Wahleithner J.A., Lund M., Bonnett H.T., Kelly A., Meeks-Wagner D.R., Peacock W.J., Dennis E.S. (1990). Chitinase, B-1,3-glucanase, osmotin and extensin are expressed in tobacco explants during flower formation. *The Plant Cell* 2: 673-684.

Newman S.M., Boynton J.E., Gillham N.W., Randolph-Anderson B.L, Johnson A.M., Harris E.H. (1990). Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: Molecular and genetic characterization of integration events. *Genetics* 126: 875-888.

Ngernprasirtsiri J., Chollet R., Kobayashi H., Sugiyama T., Akazawa T. (1989). DNA methylation and the differential expression of C4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. *J. Biol. Chem.* 264: 8241-8248.

Ngernprasirtsiri J., Kobayashi H., Akazawa T. (1988a). DNA methylation occured around lowly expressed genes of plastid DNA during tomato fruit development. *Plant Physiol.* 88: 16-20.

Ngernprasirtsiri J., Kobayashi H., Akazawa T. (1988b). DNA methylation as a mechanism of transcriptional regulation in nonphotosynthetic plastids in plant cells. *Proc. Natl. Acad. Sci. USA* 85: 4750-4754.

Nomiyama H., Fukuda M., Wakasugi S, Tsuzuki T., Shimada K. (1985). Molecular structures of mitochondrial-like-DNA sequences in human nuclear DNA. *Nucl. Acids Res.* 13: 1649-1658.

Nomiyama H., Tsuzuki T., Wakasugi S., Fukuda M., Shimada K. (1984). Interruption of a human nuclear sequence homologous to mitochondrial DNA by a member of the KpnI 1.8 kbp family. *Nucl. Acids Res.* 12: 5225-5234.

Nugent J.M., Palmer J.D. (1988). Location, identity, amount and serial entry of chloroplast DNA sequences in crucifer mitochondrial DNA's. *Curr. Genet.* 14: 501-509.

Nugent J.M., Palmer J.D. (1991). RNA-mediated transfer of the gene *cox*II from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66: 473-481.

Ohme M., Kamogashira T., Shinozaki K., Sugiura M. (1985). Structure and cotranscription of the tobacco chloroplast genes for $tRNA^{glu}$ (UUC), $tRNA^{tyr}$ (GUA) and $tRNA^{asp}$ (GUC). *Nucl. Acids Res.* 13: 1045-1056.

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. (1988a). Chloroplast gene organization deduced from the complete sequence of the liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572-574.

Ohyama K., Kohchi T., Sano T., Yamada Y. (1988b). Newly identified groups of genes in chloroplasts. *Trends Biochem. Sci.* 13: 19-22.

Ossorio P.N., Sibley L.D., Boothroyd J.C. (1991). Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. *J. Mol. Biol.* 22: 525-536.

Palmer J.D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19: 325-354.

Palmer J.D. (1985b). Evolution of chloroplast and mitochondrial DNA in plants and algae. In: Molecular evolutionary genetics pp. 131-240. MacIntyre R.J. (Ed.) Plenum, New York.

Palmer J.D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. *TIG* 6: 115-120.

Palmer J.D., Herbon L.A. (1987). Unicircular structure of the *Brassica hirta* mitochondrial genome. *Curr. Genet.* 11: 565-570.

Palmer J.D., Herbon L.A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* 28: 87-97.

Palmer J.D., Sheilds C.R. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* 307: 437-440.

Palmer J.D., Thompson W.F. (1981). Clone banks of the mung bean, pea and spinach chloroplast genomes. *Gene* 15: 21-26.

Palmer J.D., Thompson W.F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29: 537-550.

Penny D. (1989). What, if anything is a *Prochloron* ? *Nature* 337: 304-305.

Penny D., O'Kelly C.J. (1991). Seeds of a universal tree. *Nature* 350: 106-107.

Pichersky E. (1990). Nomad DNA- A model for movement and duplication of DNA sequences in plant genomes. *Plant Mol. Biol.* 15: 437-448.

Pichersky E., Logsdon (Jr.) J.M., McGrath J.M., Stasys R.A. (1991). Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location and possible mechanisms of integration. *Mol. Gen. Genet.* 225: 453-458.

Pichersky E., Tanksley S.D. (1988). Chloroplast DNA sequences integrated into a intron of a tomato nuclear gene. *Mol. Gen. Genet.* 215: 65-68.

Piechulla B., Kellman J-W., Pichersky E., Schwartz E., Forster H-H. (1991). Determination of the steady state mRNA levels of individual chlorophyll a/b binding protein genes of the tomato *cab* gene family. *Mol. Gen. Genet.* 230: 413-422.

Pring D.R., Lonsdale D.M. (1985). Molecular biology of higher plant mitochondrial DNA: *Int. Rev. Cyt.* 97: 1-46.

Raleigh E.A., Murray N.E., Revel H., Blumenthal R.M., Westaway D., Reith A.D., Rigby P.W.J., Elhai J., Hanahan H. (1988). McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. *Nucl. Acids 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.

Richards A.J. (1986). Plant breeding systems. George, Allen and Unwin Ltd., U.K., U.S.A., Aust.

Ridder R., Kunkele K-P., Osiewacz H.D. (1991). Sequence of the nuclear ATP synthase subunit 9 gene of *Podospora anserina*: lack of similarity to the mitochondrial genome. *Curr. Genet.* 20: 349-351.

Ripley L.S. (1982). Model for the participation of quasi-palindrome DNA sequences in frameshift mutations. *Proc. Natl. Acad. Sci. USA* 79: 4128-4132.

Rogers J.H. (1989). How were introns inserted into nuclear genes? *TIG* 5: 213-216.

Roth D.B., Chang X-B., Wilson J.H. (1989). Comparisons of filler DNA at immune, nonimmune and oncogenic rearrangements suggests multiple mechanisms of formation. *Mol. Cell. Biol.* 9: 3049-3057.

Roth D.B., Wilson J.H. (1986). Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol.Cell. Biol.* 6: 4295-4304.

Sainsard-Chanet A., Begel O. (1990). Insertion of an LrRNA gene fragment and of filler DNA at a mitochondrial exon-intron junction in *Podospora. Nucl. Acids Res.* 18: 779-783.

Sakamoto W., Kadowaki K., Kishimota N., Yano M., Saito A., Tano S. (1991). RFLP analysis of nuclear DNA's homologous with mitochondrial plasmid-like DNA's in cultivated rice. *Theor. Appl. Genet.* 82: 179-184.

Sambrook J., Fritsch E.F., Maniatis T. (1989). Molecular cloning. A laboratory manual (second edition). Cold Spring Harbour, New York.

Sangare A., Weil J-H., Grienenberger J-M., Fauron C., Lonsdale D.M. (1990). Localization and organization of tRNA genes on the mitochondrial genomes of fertile and male sterile lines of maize. *Mol. Gen. Genet.* 223: 224-232.

Sanger F., Nicklen S., Coulson A.R. (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463.

Scazzocchio C. (1989). Group I introns: do they only go home? *TIG* 5: 168-172.

Schaaper R.M., Danforth B.N., Glickman B.W. (1986). Mechanisms of spontaneous mutagenesis: An analysis of the spectrum of spontaneous mutations in the *Escherichia coli lac*I gene. *J. Mol. Biol.* 189: 273-284.

Schenk H.E.A., Bayer M.G., Zook D. (1988). From symbiont to organelle. *Ann. N.Y. Acad. Sci.* 503: 151-167.

Schimper A.F.W. (1883). Uber die Entwicklung der Chlorophyllkorner und Farbkorner (1 Teil). *Bot. Zeit.* 41: 105-114.

Schuster W., Brennicke A. (1987a). Plastid DNA in the mitochondrial genome of *Oenothera*: Intra- and interorganellar rearrangements involving part of the plastid ribosomal cistron. *Mol. Gen. Genet.* 210: 44-51.

Schuster W., Brennicke A. (1987b). Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera*: is genetic information transferred between organelles via RNA? *EMBO* 6: 2857-2863.

Schuster W., Brennicke A. (1988). Interorganellar sequence transfer: plant mitochondrial DNA is nuclear, is plastid, is mitochondrial. *Plant Sci.* 54: 1-10.

Scott N.S., Possingham J.V. (1980). Chloroplast DNA in expanding spinach leaves. *J. Exp. Bot.* 123: 1081-1092.

Sederoff R.R., Ronald P., Bedinger P., Rivin C., Walbot V., Bland M., Levings III C.S. (1986). Maize mitochondrial plasmid S-1 sequences share homology with chloroplast gene *psb*A. *Genetics* : 469-482.

Shih M-C., Lazar G., Goodman H.M. (1986). Evidence in favour of the symbiotic origin of chloroplasts: Primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenase. *Cell* 47: 73-80.

Shikanai T., Yang Z-Q., YamadaY. (1987). Properties of the circular plasmid-like DNA B1 from mitochondria of cytoplasmic male-sterile rice. *Plant Cell Physiol.* 28: 1243-1251.

Shimada H., Sugiura M. (1991). Fine structural features of the chloroplast genome: Comparisons of sequenced chloroplast genomes. *Nucl. Acids Res.* 19: 983-995.

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. (1986). The complete sequence of the tobacco chloroplast genome: its organization and expression. *EMBO* 5: 2043-2049.

Siculella L., Palmer J.D. (1988). Physical and gene organization of mitochondrial DNA in fertile and male sterile sunflower. CMS-associated alterations in structure and transcription of the *atp*A gene. *Nucl. Acids Res.* 16: 3787-3799.

Simmonds N.W. (1971). The breeding system of *Chenopodium quinoa*. *Heredity.* 27: 72-82.

Simmonds N.W. (1976). Evolution of crop plants. Longman Group Ltd., New York.

Small I., Suffolk R., Leaver C.J. (1989). Evolution of plant mitochondrial genomes *via* substoichometric intermediates. *Cell* 58: 69-71.

Smillie R.M., Scott N.S. (1969). Organelle biosynthesis: the chloroplast. In: Progress in molecular biology, vol.I.pp 136-202. Springer, Berlin, Hedielberg, New York.

Smith A.G., Pring D.R. (1987). Nucleotide sequence and molecular characterization of a maize mitochondrial-plasmid-like DNA. *Curr. Genet.* 12: 617-623.

Smooker P.M., Choli T., Subramanian A.R. (1990). Ribosomal protein L35: identification in spinach chloroplasts and isolation of a cDNA clone encoding its cytoplasmic precursor. *Biochemistry* 29: 9733-9736.

Smooker P.M., Kruft V., Subramanian A.R. (1990). A ribosomal protein is encoded in the chloroplast DNA in a lower plant but in the nucleus in angiosperms. Isolation of the spinach L21 protein and cDNA clone with a transit and unusual repeated sequence. *J. Biol. Chem.* 265: 16699-16703.

Sorge J.A. (1988). Bacteriophage lambda cloning vectors and their uses. R.L. Rodrignez, D.T. Denhardt (ed.) Butterworth.

Southern E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503.

Sporlein B., Streubel M., Dahlfeld G., Westhoff P., Koop H.V. (1991). PEG-mediated plasmid transformation: a new system for transient gene expression assays in chloroplasts. *Theor. Appl. Genet.* 82: 717-722.

Steele Scott N., Timmis J.N. (1984). Homologies between nuclear and plastid DNA in spinach. *Theor. Appl. Genet.* 67: 279-288.

Steele Scott N., Timmis J.N., Ayliffe M.A. (1991). Chloroplast genes in higher plants. *Aust. J. Plant Physiol.* 18: 533-541.

Stern D.B., Gruissem W. (1987). Control of plastid gene expression: 3' inverted repeats act as a mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* 51: 1145-1157.

Stern D.B., Lonsdale D.M. (1982). Mitochondrial and chloroplast genomes of maize have a 12-kb DNA sequence in common. *Nature* 299: 698-702.

Stern D.B., Palmer J.D. (1984). Extensive and widespread homologies between mitochondrial and chloroplast DNA in plants. *Proc. Natl. Acad. Sci. USA* 81: 1946-1950.

Stern D.B., Palmer J.D. (1986). Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping and locations of transposed sequences. *Nucl. Acids Res.* 14: 5651-5666.

Sugiura M. (1989). The chloroplast chromosome in land plants. *Ann. Rev. Cell Biol.* 5: 51-70.

Sugiura M., Shinozaki K., Zaita N., Kusuda M., Kumano M. (1986). Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genes. *Plant Sci.* 44: 211-216.

Suyama Y. (1986). Two dimensional polyacrylamide gel electrophoresis analysis of *Tetrahymena* mitochondrial tRNA. *Curr. Genet.* 10: 411-420.

Tanaka M., Wakasugi T., Sugita M., Shinozaki K., Sugiura M. (1986). Genes for eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): similarity to the S10 and spc operons of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 83: 6030-6034.

Taylor F.J.R. (1988). An overview of the status of evolutionary cell symbiosis theories. *Ann N.Y. Acad. Sci.* 503: 1-17.

Thorness P.E., Fox T.D. (1990). Escape of DNA from the mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* 346: 376-379.

Thuring R.W.J., Sanders J.P.M., Borst P. (1975). A freeze squeeze method for the recovery of long DNA from agarose gels. *Anal. Biochem.* 66: 213-220.

Timmis J.N., Steele Scott N. (1983). Sequence homology between spinach nuclear and chloroplast genomes. *Nature* 305: 65-67.

Trinh T.Q., Sinden R.R. (1991). Preferential DNA secondary structure mutagenesis in the lagging strand of replication of *E. coli*. *Nature* 352: 544-547.

Tsuzuki T., Nomiyama H., Setoyama C., Maeda S., Shimada K. (1983). Presence of mitochondrial-DNA-like sequences in the human nuclear DNA. *Gene* 25: 223-229.

Turner S., Burger-Wiersman T., Giovannoni S.J., Mur L.R., Pace N.R. (1989). The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature* 337: 380-382.

Turpen T., Garger S.J., David-Marks M., Grill L. K. (1987). Molecular cloning and physical characterization of a *Brassica* linear mitochondrial plasmid. *Mol. Gen. Genet.* 209: 227-229.

van den Boogaart P., Samallo J., Agsteribbe E. (1982) Similar genes for a mitochondrial ATPase subunit in the nuclear genome of *Neurospora crassa. Nature* 298: 187-189.

van Loon L.C., Trewavas A., Chapman K.S.R. (1975). Phosphorylation of chromatin associated proteins in *Lemna* and *Hordeum. Plant Physiol.* 55: 288-292.

Vera A., Tomas R., Martin M., Sabater B. (1990). Apparent expression of the small single copy cpDNA region in senescent chloroplasts. *Plant Sci.* 72: 63-67.

Vestweber D., Schatz G. (1989). DNA-protein conjugates can enter mitochondria *via* the protein import pathway. *Nature* 338: 170-172.

Vierny C., Keller A.M., Begel O., Belcour L. (1982). A sequence of mitochondrial DNA associated with the onset of senescence in a fungus. *Nature* 297: 157-159.

Wakasugi S., Nomiyama H., Fukuda M., Tsuzuki T., Shimada K. (1985). Insertion of a long KpnI family member within a mitchondrial-DNA-like sequence present in the human nuclear genome. *Gene* 36: 281-288.

Ward B.L., Anderson R.S., Bendich A.J. (1981). The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell* 25: 793-803.

Weaver D.T., DePamphilis M.L. (1984). The role of palindromic and nonpalindromic sequences in arresting DNA synthesis *in vitro* and *in vivo. J. Mol. Biol.* 180: 961-986.

Weeden N.F. (1981). Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J. Mol. Evol.* 17: 133-139.

Westhoff P., Alt J., Nelson N., Bottomley W., Bunemann H., Herrman R.G. (1983). Genes and transcripts for the P700 chlorophyll a apoprotein and subunit 2 of the photosystem reaction center complex from spinach thylakoid membranes. *Plant Mol. Biol.* 2: 95-107.

Whisson D.L., Steele Scott N. (1985). Nuclear and mitochondrial DNA have sequence homology with a chloroplast gene. *Plant Mol. Biol.* 4: 267-273.

Wimpee C.F., Wrobel R.L., Garvin D.K. (1991). A divergent plastid genome in *Conopholis americana*, a chlorophyllous parasitic plant. *Plant Mol. Biol.* 17: 161-166.

Wintz H., Chen H-C., Pillay D.T.N. (1988a). Presence of a chloroplast-like elongator $tRNA^{met}$ gene in the mitochondrial genomes of soybean and *Arabidopsis thaliana. Curr. Genet.* 13: 255-260.

Wintz H., Grienenberger J.M., Weil J-H., Lonsdale D.M. (1988b). Location and nucleotide sequence of two tRNA genes and a tRNA pseudo-gene in the maize mitochondrial genome: evidence for the transcription of a chloroplast gene in the mitochondria. *Curr. Genet.* 13: 247-

Woodcock D.M., Crowther P.J., Diver W.P., Graham M., Bateman C., Baker D.J., Smith S.S. (1988). RglB facilitated cloning of highly methylated eucaryotic DNA: the human L1 transposon, plant DNA, and DNA methylated in vitro with human DNA methyltransferase. *Nucl. Acids Res.* 16: 4465-4483.

Woodcock D.M., Crowther P.J., Doherty J., Jefferson S., DeCruz E., Noyer-Weidner M.,Smith S.S., Michael M.Z., Graham M.W. (1989). Quantitative evaluation of *Escherichia coli* strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucl. Acids Res.* 17: 3469-3478.

Wright R.M. (1982). Are mitochondrial structural genes selectively amplified during senescence in *Podospora anserina* ? *Cell* 29: 505-515.

Wright R.M., Cummings D.J. (1983). Integration of mitochondrial gene sequences within the nuclear genome during senescence in a fungus. *Nature* 302: 86-88.

Xu M-Q., Kathe S.D., Goodrich-Blair H., Nierzwicki-Baeur S.A., Shub D.A. (1990). Bacterial origins of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* 250: 1566-1569.

Yao W.B., Meng B.Y., Tanaka M., Sugiura M. (1989). An additional promoter within the protein-coding region of the *psb*D-*psb*C gene cluster in tobacco chloroplast DNA. *Nucl. Acids Res.* 17: 9583-9591.

Zurawski G., Perrot B., Bottomley W., Whitfeld P.R. (1981). The structure of the gene for the large subunit of the enzyme ribulose-1,5-bisphosphate carboxylase from spinach. *Nucl. Acids Res.* 9: 3251-3270.