THE MOLECULAR BASIS OF CYTOPLASMIC MALE STERILITY
IN HELIANTHUS ANNUUS (SUNFLOWER)

BY

HILARY KATHLEEN LAVER

DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH
1989
Table of contents

Table of Contents
Acknowledgements
Abstract
Abbreviations

Chapter One
INTRODUCTION
1.1 The Structure and Function of Mitochondria 1
1.2 Mitochondrial Biogenesis- the Nuclear Contribution 4
   1.2.1 The Endosymbiont Theory 5
   1.2.2 Protein Import into mitochondria 6
1.3 Unique Features of Plant Mitochondria 7
1.4 Organellar Genomes - Organisation and Recombination 9
   1.4.1 The Structure and Organisation of Mitochondrial Genomes 9
   1.4.2 Higher Plant Mitochondrial Genomes 11
   1.4.3 Mechanisms of Genome Recombination 17
   1.4.4 Chloroplast Genomes 18

1.5 The Coding Capacity of Plant Mitochondrial Genomes 21
   1.5.1 Mitochondrial Ribosomal RNA Genes 21
   1.5.2 Mitochondrial Transfer RNA Genes 22
   1.5.3 Mitochondrial Protein Coding Genes 23

1.6 Plasmid-like DNAs associated with Plant Mitochondria 27
1.7 Promiscuous DNA 27

1.8 Expression of Mitochondrial Genes 28
   1.8.1 Mammalian mitochondrial gene expression 28
   1.8.2 Yeast mitochondrial gene expression 29
      i. Transcript Initiation 29
      ii Transcript Termination 30
      iii Transcript Processing 30
   1.8.3 Plant Mitochondrial Gene Expression 31
      i Transcript Initiation 31
      ii Transcript Termination 32
      iii Transcript Processing 33
      iv Cotranscription 34
   1.8.4 Transcripts with Unassigned Functions in Plant Mitochondria 34
   1.8.5 RNA editing in Plant Mitochondria 35
   1.8.6 Translation of Plant Mitochondrial Genes 36

1.9 Phenotypic Mutations Associated with Mitochondrial DNA Alterations 38
   1.9.1 Non-chromosomal Stripe Mutations in Maize 38
   1.9.2 Cytoplasmic Male Sterility 38
   1.9.3 The Use of CMS in Plant Breeding 39
   1.9.4 CMS in maize
      i CMS-T 40
      ii CMS-C 41
      iii CMS-S 42
   1.9.5 CMS in Petunia 43

1.10 Cytoplasmic Male Sterility in Sunflower 44
   1.10.1 Ultrastructural Observations of Microsporogenesis in Sunflower 45
   1.10.2 Molecular Analysis of CMS in Sunflower 47
   1.10.3 Aims of this thesis project 47
## Chapter Two MATERIALS AND METHODS

### 2.1 Materials

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1</td>
<td>Seed material</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Chemicals</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Enzymes</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Radioisotopes</td>
</tr>
<tr>
<td>2.1.5</td>
<td>Film and Intensifying screens</td>
</tr>
<tr>
<td>2.1.6</td>
<td>Centrifuge Equipment</td>
</tr>
<tr>
<td>2.1.7</td>
<td>Mitochondrial DNA clones</td>
</tr>
<tr>
<td>i</td>
<td>Maize mtDNA clones</td>
</tr>
<tr>
<td>ii</td>
<td><em>Oenothera</em> mtDNA clones</td>
</tr>
<tr>
<td>2.1.8</td>
<td>Bacterial Host Strains and cloning vectors</td>
</tr>
<tr>
<td>2.1.9</td>
<td>Other materials</td>
</tr>
</tbody>
</table>

### 2.2 Methods

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1</td>
<td>Growth of plant Material</td>
</tr>
<tr>
<td>i</td>
<td>Growth of Etiolated Seedlings</td>
</tr>
<tr>
<td>ii</td>
<td>Growth under diurnal light/ temperature conditions</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Purification of Sunflower Mitochondria</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Isolation of Mitochondrial DNA</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Electrophoresis of Sunflower Mitochondrial DNA</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Mitochondrial DNA Blot Hybridisation</td>
</tr>
<tr>
<td>2.2.6</td>
<td>Preparation of Sunflower Mitochondrial RNA</td>
</tr>
<tr>
<td>2.2.7</td>
<td>Preparation of Total RNA from Sunflower</td>
</tr>
<tr>
<td>2.2.8</td>
<td>RNA Gel Electrophoresis and Hybridisation</td>
</tr>
<tr>
<td>2.2.9</td>
<td>Preparation of Chloroplast DNA</td>
</tr>
<tr>
<td>2.2.10</td>
<td>Cosmid Cloning of Sunflower Mitochondrial DNA</td>
</tr>
<tr>
<td>i</td>
<td>Preparation of vector DNA</td>
</tr>
<tr>
<td>ii</td>
<td>Preparation of Insert DNA</td>
</tr>
<tr>
<td>iii</td>
<td>Ligation and Packaging</td>
</tr>
<tr>
<td>iv</td>
<td>Cosmid Transformation</td>
</tr>
<tr>
<td>2.1.11</td>
<td>M13 Cloning</td>
</tr>
<tr>
<td>2.1.12</td>
<td>Plasmid Cloning</td>
</tr>
<tr>
<td>2.1.13</td>
<td>Preparation of Cloned DNA</td>
</tr>
<tr>
<td>i</td>
<td>Preparation of M13 Single-stranded DNA</td>
</tr>
<tr>
<td>ii</td>
<td>Preparation of Plasmid DNA</td>
</tr>
<tr>
<td>iii</td>
<td>Cosmid DNA Preparation</td>
</tr>
<tr>
<td>2.2.14</td>
<td>DEAE Gel Purification</td>
</tr>
<tr>
<td>2.2.15</td>
<td>Preparation of Radioactive Probes</td>
</tr>
<tr>
<td>i</td>
<td>TCA precipitation</td>
</tr>
<tr>
<td>ii</td>
<td>Nick translation</td>
</tr>
<tr>
<td>iii</td>
<td>Primer Extension</td>
</tr>
<tr>
<td>iv</td>
<td>Random Probe Priming</td>
</tr>
<tr>
<td>v</td>
<td>Oligo-nucleotide end-labeling</td>
</tr>
<tr>
<td>vi</td>
<td>Mitochondrial DNA end-labeling</td>
</tr>
<tr>
<td>2.2.16</td>
<td>DNA Sequencing</td>
</tr>
<tr>
<td>2.2.17</td>
<td>Mitochondrial Protein Synthesis</td>
</tr>
<tr>
<td>i</td>
<td>Incorporation of 35S-methionine</td>
</tr>
<tr>
<td>ii</td>
<td>Protein Gel Electrophoresis</td>
</tr>
</tbody>
</table>

Chapter Three MITOCHONDRIAL GENES, GENOMES AND MOLECULAR REARRANGEMENT IN SUNFLOWER

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction and Rationale</td>
</tr>
<tr>
<td>3.2</td>
<td>Isolation of Native Mitochondrial DNA from Fertile and Sterile Sunflower</td>
</tr>
</tbody>
</table>
3.3 Estimation of Mitochondrial Genome Size by Digestion with Restriction Endonucleases
3.4 Investigation of the Mitochondrial Genome Organisation in Fertile and Sterile Sunflower by Mt DNA Hybridisation Analysis
3.5 The Sunflower Mitochondrial Genome Contains Substoichiometric Homologies to the atpA Gene
3.6 The atp9 Gene May Be Duplicated in the Mitochondrial Genome of Sunflower
3.7 The coxl gene in Sunflower Contains an Intron
3.8 Homology to Sequences Encoding Subunits 1, 3 and 5 of the NADH dehydrogenase are Present in Sunflower MtDNA
3.9 Sunflower MtDNA Contains Linked 18S and 5S Ribosomal Genes
3.10 The rps12 gene is Encoded by the Sunflower Mitochondrial Genome
3.11 The Use of Cosmid Clones to Investigate Rearrangement in the Mitochondrial Genomes of Fertile and Sterile Sunflower
3.12 Construction of Cosmid Libraries from Fertile and Sterile Sunflower following Partial Digestion of MtDNA
3.13 Detailed Mapping Around atpA Indicates a Divergence in the 3' Region of the Gene in Sterile Sunflower
3.14 Mapping of the cob Gene Indicates that it is Closely Linked to atpA in the Mitochondrial Genome of Sterile Sunflower
3.15 Rearrangement of the Mitochondrial Genome of Sterile Sunflower Involves Both Sequence Inversion and Insertion
3.16 A Novel Mitochondrial Plasmid is Amplified in Restorer Sunflower
3.17 Conclusions regarding Mitochondrial Genome Organisation in Fertile and Sterile Sunflower
3.18 Mitochondrial Genome Heterogeneity in Higher Plants

Chapter Four SEQUENCE ANALYSIS OF THE ATPA GENE FROM FERTILE AND STERILE SUNFLOWER
4.1 Aims and Rationale
4.2 Strategy for Sequencing the atpA gene from Fertile and Sterile Sunflower
4.3 Nucleotide Sequence Analysis of the atpA gene from Fertile and Sterile Sunflower
4.4 Comparison of the Predicted Amino Acid Sequence of the Sunflower Mitochondrial atpA gene with Other F1 ATPase α-subunit proteins
4.5 Sequence Divergence Occurs 3' to the atpA genes in Fertile and Sterile Sunflower
4.6 Open Reading Frames Downstream of the atpA gene in Fertile and Sterile Sunflower
Chapter Five

TRANSCRIPTION OF THE MITOCHONDRIAL GENOME FROM FERTILE AND STERILE SUNFLOWER

5.1 Introduction and Aims
5.2 AtpA Transcription is Altered in Sterile Sunflower
5.3 Cob Transcription is Unaltered Between Fertile and Sterile Sunflower
5.4 Transcripts for All Other Genes Examined are Identical in Fertile and Sterile Sunflower Mitochondria
5.5 Transcription During Flower Development in Fertile and Sterile Sunflower
5.6 Quantitative Variation in Total RNA and MtRNA in Fertile and Sterile Anthers
5.7 Mitochondrial Transcripts are Most Abundant During Early Flower Development
5.8 Analysis of Mitochondrial DNA During Flower Development
5.9 A Comparison of Nuclear and Mitochondrial Transcription in Fertile and Sterile Sunflower During Fertile and Sterile Sunflower Development
5.10 Conclusions

Chapter Six

MITOCHONDRIAL PROTEIN SYNTHESIS IN SUNFLOWER

6.1 Introduction
6.2 The Problem of Bacterial Contamination
6.3 The Effect of Energy Source on Protein Synthesis by Isolated Sunflower Mitochondria
6.4 A Comparison of Polypeptides Synthesized by Mitochondria from Fertile and Sterile Sunflower
6.5 Conclusions

Chapter Seven

DISCUSSION

7.1 Sunflower- An Ideal System for Studying Cytoplasmic Male Sterility
7.2 Variant Expression of the Mitochondrial Genome of Sterile Sunflower
7.3 Rearrangements in the Sunflower Mitochondrial Genome - Sublimons and Speculation
7.4 Short Tandem repeats and Mitochondrial Genome Rearrangement
7.5 The 'Plasmid-like DNA in Restorer Sunflower May be an Indication of a Nuclear Influence on mtDNA Replication
7.6 Synthesis of Variant Mitochondrial Proteins in Sunflower
7.7 Molecular Events in the Cytoplasm During Pollen Formation
7.8 The Molecular Basis of Cytoplasmic Male Sterility in Sunflower
7.9 Proposals for Future Work
  7.9.1 Sequence and Transcript Mapping
  7.9.2 Anther Development and Transcription
  7.9.3 Mitochondrial Protein Purification and Functional Characterisation of the Variant 15kDa Polypeptide
  7.9.4 Biochemical Characterisation of Sunflower Mitochondria
7.10 The Changing Role of Mitochondria During Plant Development
7.11 Genetic Engineering of Plant Mitochondria- Some Future Prospects
7.12 Conclusion

REFERENCES
Acknowledgements

I am sincerely indebted to a number of people for their help over the last three years.

To Chris Leaver, for his guidance, support and the volumes of caffeine which have fuelled the completion of this thesis. I am also grateful for the helpful comments made by Dr Tony Trewevas and Dr Steve Smith during the course of this project. Financial Support was kindly provided by Rhône-Poulenc Agrochimie of France and I would like to thank Professor Georges Freyssinet, in particular, for assistance in supplying all seed material and for his interest in the research. Also I wish to thank Professor Hugh Dickinson of Reading University for stimulating discussions and for the time and effort spent on microscope examinations. For advice on the technicalities associated with cosmid cloning, I am indebted to Dr David Lonsdale of the Plant Breeding Institute, Cambridge and Dr Yukimoto Iwasaki. Andrew Liddell provided invaluable assistance during the mitochondrial protein experiments. I shall be eternally grateful to Messers Roger and Joe McChuskey for managing to rescue my flowering Helianthus petiolaris from force ten gales and low flying greenhouses.

I also owe great deal to all my friend and colleagues in Edinburgh, past and present, but especially C.J (see!), Brenda, Daisy, Ian, Joe, Jacques and Mamie, all of whom I shall miss. Also Wizzie, for hangin' in there, and Julie and Barbara for being there. I appreciate the support and tolerance of assorted members of my family, especially my mother, simply for understanding, and my father for accepting a perennial student as a daughter.

I dedicate this thesis to Mike, for having confidence in me, for being so patient, but mostly for making me laugh.
Abstract

A comparative investigation of the organisation and expression of the mitochondrial genome in fertile and cytoplasmic male sterile (CMS) Helianthus annuus (sunflower) has been undertaken. The aim of the work described was to identify mitochondrial genome variation between the two phenotypes which correlates with the expression of male sterility. Several experimental approaches were used, aimed at defining how specific mitochondrial genome rearrangements found in sterile sunflower relate to altered patterns of transcription and translation within this cytoplasm.

Restriction endonuclease digestion 'profiles' of sunflower mitochondrial DNA (mtDNA) leads to an estimate of ca. 300kb for the size of the mitochondrial genome in both cytoplasms. The similarity of these profiles indicates that there is relatively little variation in the mitochondrial genome organisation between fertile and sterile sunflower. Hybridisation of specific mitochondrial gene probes to mtDNA from both cytoplasms indicates that the coding regions of most genes analysed are identical. However, a rearrangement was identified in the sterile sunflower genome, in the vicinity of the atpA locus which encodes the alpha subunit of the F₁ ATP synthase complex of the inner mitochondrial membrane.

Restriction endonuclease mapping has been used to show that a sequence homologous to the 3' flank of the atpA gene is repeated approximately 13kb downstream of the intact gene in fertile sunflower. Sequence analysis has shown that this repeated sequence extends for 200-300 bases and is indirect with respect to the atpA gene. Homologous recombination across this repeat is proposed to have mediated an inversion of the region between these repeats to generate the arrangement found in the mitochondrial genome of the sterile sunflower. In addition, a sequence insertion of ca. 2.3kb has occurred at the 5' end of this inversion in the sterile sunflower genome. The predicted protein sequence of the atpA open reading frame (ORF) is identical between fertile and sterile sunflower but two novel ORFs immediately downstream of the atpA gene in sterile sunflower have been generated as a consequence of the rearrangements.

An investigation of expression of the atpA gene indicates that an abundant transcript of 1.9kb synthesised in fertile sunflower is replaced by an extended 3kb transcript in sterile sunflower. In addition, two other transcripts, homologous to the 3' flanking region of the atpA gene are found. A 1.3kb transcript in fertile sunflower correlates with the presence of potential coding sequences downstream of atpA in this genome. A larger 1.7kb transcript in sterile sunflower correlates with the finding of the two novel ORFs in the rearranged region downstream of atpA in this genome.

The spectrum of proteins synthesised by mitochondria isolated from fertile and sterile sunflower have been compared by 'in organello' incorporation of 35S-labelled methionine. This has demonstrated the presence of a ca. 15kDa polypeptide synthesised in mitochondria from sterile sunflower, but not those isolated from fertile sunflower. A model is presented, based on the observations made, for a possible mechanism by which the molecular variation found could be causally associated with the CMS phenotype.

In addition, a previously unreported 5kb plasmid has been found in all sunflower lines examined, but which is greatly amplified in copy number in the restorer line 843R. Interestingly, this plasmid shows homology to the region of the main mitochondrial genome which has become rearranged between fertile (842B) and sterile (842A) sunflower. A possible involvement of the 5kb plasmid in the restoration of fertility is discussed.
**Abbreviations**

3' 3' carbon of a nucleotide or nucleoside, hydroxy terminus of a nucleic acid backbone, towards the 3' carbon

5' 5' carbon of a nucleotide or nucleoside, hydroxy terminus of a nucleic acid backbone, towards the 5' carbon

% percentage

α alpha

ANT adenine nucleotide translocator

Ant-A antimycin A

APS ammonium persulphate

ATP adenosine triphosphate

atpA, atp6, atp9 mitochondrial nucleotide sequences encoding subunits α, 6 and 9 respectively of the F1-F0-ATPase

ATPA α subunit of the F1ATPase

ATP9 subunit 9 of the F0-ATPase

β beta

BDH British Drug Houses

bisacrylamide N, N'-methylenebisacrylamide

BM Boehringer Mannheim

bp base pairs

Bq bequerels

BRL Bethesda Research Laboratories

BSA bovine serum albumin

°C degrees centigrade

CAM Crassulacean Acid Metabolism

CMS cytoplasmic male sterility

CMS-T, -S, -C CMS maize types Texas, USDA and Charrua

CN cyanide

cob mitochondrial gene encoding apocytochrome b

COB apocytochrome b

cox I, II, III mitochondrial genes encoding subunits 1, 2 and 3 of the cytochrome oxidase complex (complex IV)

COXI cytochrome oxidase subunit I

cpDNA chloroplast DNA

CsCl cesium chloride

dDCD N,N'-dicyclohexylcarbodiimide

DEAE diethylaminoethyl

DF Difco Laboratories

dH2O distilled water

DNA deoxyribonucleic acid

dpm disintegrations per minute

ds double-stranded

DTT dithiothreitol

EDTA ethylenediaminetraacetic acid

EGTA ethyleneglycol-bis-N,N', N', N'-tetraacetic acid

FAD flavine adenine dinulceotide

FADH flavine adenine dinulceotide reduced form

F1-ATPase F1 portion of the F1-F0 ATP synthetase of the inner mitochondrial membrane

F0-ATPase F0 portion of the F1-F0 ATP synthetase
\(\gamma\)  gamma

g  gram

GTP  guanosine 5'-triphosphate

HCl  hydrochloric acid

IPTG  isopropyl \(\beta\)-C-thiogalactoside

kb  kilobase pair(s)

kDa  kilodalton

l  litre

LB  Luria broth (described in section 2.2.13)

LiCl  lithium chloride

Ltd  limited

M  molar

mA  milliamperes

MES  2-[N-morpholinol ethanesulphonic acid

mg  milligram

min.  minute

ml  millilitre

mM  millimolar

MOPS  3-[N-morpholinol propanesulphonic acid

MPP  mitochondrial processing peptidase

mRNA  messenger RNA

mtDNA  mitochondrial DNA

N\(_2\)  liquid nitrogen

NaAc  sodium acetate

NaCl  sodium chloride

NaOH  sodium hydroxide

nad1  mitochondrial reading frame encoding a protein showing homology to subunit I of the NADH ubiquinone oxidoreductase complex (complex I)

nad3, nad5  mitochondrial genes encoding subunits 3 and 5 of complex I

NAD  nicotinamide adenine dinucleotide

NADH  nicotinamide adenine dinucleotide (reduced from)

NADPH  nicotinamide adenine dinucleotide phosphate (reduced form)

NCS  non-chromosomal stripe

NDS  napthalene-1,5-disulphonic acid

nm  nanometre

ng  nanogram

N-terminal  amino terminal

N-type  normal, fertile (maize)

O\(_2\)  molecular oxygen

ORF  open reading frame

PAS  4-aminosalicylate

PEG  polyethylene glycol

PEP  processing enhancer protein

pH  negative log of hydrogen ion concentration

plc  public limited company

PVP  polyvinylpyrrolidone

Rf, rf  nuclear restorer gene

RNA  ribonucleic acid

RNAse  ribonulease

rpm  revolutions per minute

rps12, 13  mitochondrial genes encoding ribosomal proteins 12 and 13
rRNA  ribosomal RNA
Rt    rotenone
S     Svedberg units
SCC   Sigma Chemical Company
sdH₂O sterile distilled water
SDS   sodium dodecyl sulphate
SHAM  salicylhydroxamic acid
ss    single-stranded
SSC   standard saline citrate (defined in section 2.2.5)
S-type CMS-S type sterile (maize)
TBE   tris borate electrophoresis buffer (described in 2.2.4)
TCA   tricarboxylic acid cycle
TE    10mM Tris-HCl, 1mM EDTA
TEMED N, N, N', N'-Tetramethylenediamine
TIR   terminal inverted repeat
TNS   triisopropyl napthalene sulphonate
Tris  Tris-hydroxymethylaminomethane
Tris-HCl Tris-hydroxymethylaminomethane to pH with HCl
tRNA  transfer RNA
μCi   microcurie
μg    microgram
μl    microlitre
μM    micromolar
USA   United States of America
UV    ultra-violet
v/v   volume per volume (as a percentage)
w/v   weight per volume (as a percentage)
xg    average maximum relative gravitational force
X-Gal 5-bromo-4-chloro-3-indolyl-β-D-galactosidase
CHAPTER ONE

INTRODUCTION

The development of a higher plant is a complex process of cell division and specialisation which involves interactions between environmental cues and differential gene expression. From an early stage in plant development, cellular function and differentiation depends upon energy and metabolites provided by the mitochondria. This includes chloroplast biogenesis and the acquisition of photosynthetic competence which, at least initially, is dependent upon mitochondrial function. It is not known how mitochondrial biogenesis itself is controlled, nor how functions within mitochondria are regulated to meet the changing energy and metabolic demands of cells during differentiation.

In plant cells, three distinct genetic systems are present in the nucleus, the mitochondrion and the chloroplast. The mitochondrial and chloroplast genomes contain only a fraction of the genetic information necessary to specify these organelles. The vast majority of proteins required for the structure and metabolic activity of mitochondria and chloroplasts are nuclearly encoded, synthesized on cytosolic ribosomes and then imported into the organelles. Coordination of gene expression and protein synthesis within, and between, these cellular compartments must occur to produce functionally integrated cells. At the molecular level, we are only beginning to understand the nature of nuclear and mitochondrial interactions which are essential for growth and development in plants and other organisms.

1.1 The Structure and Function of Mitochondria

The term mitochondria is derived from two greek words, 'mitos' meaning thread-like and 'chondrion' meaning granule, and is a reflection of some of the earliest microscopic examinations of subcellular structure made in the late 19th century (Benda 1898). Today, improved electron microscopy techniques show the most common shape for mitochondria is rod-like, with hemispherical ends and dimensions approximating those of an *Escherichia coli* bacterium, about 0.5μm in diameter and 2μm in length. However, examination of serial sections through cells of chlorophytic algae such as *Chlamydomonas* and *Chlorella* and the ascomycete fungus *Saccharomyces cerevisiae* reveal that the 'individual' mitochondrial particles which appear in thin sections actually comprise a single reticulate network (Gunning and Steer 1975). Similarly, time lapse photography of higher plant cells show that, far from being discrete static structures, mitochondria are extremely dynamic organelles capable of profound changes in size, form and location. Mitochondria within plant cells appear to move along streaming cytoplasmic channels and can meet, fuse and then break up into smaller parts which subsequently disperse (Honda *et al.* 1966). The fission/fusion cycles and
contractile and expansive shape changes shown by mitochondria have been the subject of many descriptive studies, but as yet there is no complete explanation as to the cause and purpose of these movements.

The conserved function of mitochondria in all organisms is in cellular respiration, involving the complete oxidation of carbohydrate to carbon dioxide and water. Energy derived from respiration and oxidation of other organic compounds is used by the mitochondria to synthesize ATP in a process known as oxidative phosphorylation. Subsequent hydrolysis of ATP then provides energy to be utilised by cells in the biosynthetic and osmotic reactions which underlie growth and differentiation. Other key processes occur within mitochondria such as β–oxidation of fatty acids, metabolism of amino acid carbon skeletons during nitrate assimilation and the biosynthesis of nucleic acids, proteins and porphyrins (Douce 1985).

Mitochondria have a double membrane system and two separate membrane bound compartments. A smooth outer membrane surrounds and encloses the intermembrane space, an inner membrane and a completely internal protein-rich inner space called the matrix. The matrix contains the enzymes of the TCA cycle and other soluble components of mitochondrial metabolism. The polypeptide content of the matrix accounts for over 60% of the total organelle protein (Douce, 1985). In contrast, relatively few proteins are found in the intermembrane space. The outer mitochondrial membrane is not selectively permeable and allows hydrophyllic solutes of up to several thousand daltons to pass into the inner membrane space. As a proportion of the total organelle protein, the outer membrane represents only 7% and most of this (over 70%) is due to the channel-forming protein porin found in plant, fungal and animal mitochondria (Zalman et al. 1980).

The inner mitochondrial membrane comprises 30% of the total organelle protein and is one of the most complex of all biological membranes. Numerous invaginations called cristae, give the inner membrane a highly folded topology and a surface area which is almost twice that of the entire organelle. The inner membrane is highly specialised for the processes of oxidative phosphorylation and energy-linked ion translocation and contains the redox complexes, flavoproteins, iron-sulphur proteins, ubiquinones and cytochromes which constitute the respiratory chain (see Figure 1.1). These complexes catalyse the reduction of molecular oxygen (O₂) by the coenzymes nicotinamide adenine dinucleotide (NAD) and flavin adenine dinucleotide (FAD). The reduced forms of these coenzymes are generated by the oxidation of pyruvate during the tricarboxylic acid (TCA) cycle in the mitochondrial matrix and by glycolysis in the cytosol. Electrons are transferred from the reduced cofactors along the respiratory chain and at three sites, complexes I, III and IV translocate protons (H⁺ ions) from the matrix, across the inner membrane to the inner membrane space. This transfer results in an accumulation
The energy-transducing components of the higher plant respiratory chain are shown in a diagrammatic representation of their sequential arrangement in the oxidation pathways. The complexes are: I-NADH:ubiquinone oxidoreductase, II-Succinate ubiquinone oxidoreductase, III-Cytochrome bc1, IV-Cytochrome oxidase, V-F_{1}F_{0}ATPase, VI-adenine nucleotide translocator (ANT), VII-external NADH dehydrogenase and Alt-alternative oxidase. NADH oxidation is coupled to three energy-transducing sites (H^{+} production) through complex I and to two sites through the external dehydrogenase. The alternative pathway bypasses the coupling sites. Inhibitors are Rotenone (Rt), antimycin A (Ant-A) and salicylhydroxamic acid (S) (also known as SHAM).
of protons on one side of the inner membrane and energy storage in the form of a proton gradient. This energy is released when protons are returned to the matrix via the $F_1F_0$ ATP synthetase complex (ATPase, complex V) coupled to the synthesis of ATP from adenosine diphosphate (ADP) and inorganic phosphate ($P_i$).

The selectively permeable nature of the inner membrane acts as an important regulator of substrate availability to the enzymes of the TCA cycle and the ATPase complex. In addition, specific exchange mechanisms located within the inner membrane allow for the transport of selected anions into the matrix (Douce 1985). These include the adenine nucleotide translocator for the exchange of ADP and ATP (complex VI, Figure 1.1) which is directly linked to respiratory metabolism and occurs in mitochondria from all tissues. Other translocators, such as the dicarboxylic and tricarboxylic carriers involved in the exchange of compounds such as succinate, malate and citrate, are correlated with specialisation of tissues for specific metabolic functions (DeSantis et al. 1975). The inner mitochondrial membrane also contains specialised carriers for the export of substrates such as glutamate and aspartate from the mitochondria for use in biosynthetic reactions elsewhere. Obviously, the function of mitochondria, in the provision of ATP and substrates for general cellular use, requires a multitude of reactions, which must themselves be energetically coordinated and involve hundreds of proteins. Current estimates suggest that less than 10% of the total protein content of mitochondria is encoded by the mitochondrial DNA (mtDNA).

1.2 Mitochondrial biogenesis - the nuclear contribution

Mitochondria contain not only their own genomes but also the endogenous systems necessary for replication, transcription and translation of this genetic information. Isolated mitochondria will synthesize proteins in vitro in the presence of $^{35}$S-methionine. The labelled polypeptides produced by the incorporation of the radioactive methionine can then be fractionated by SDS-polyacrylamide gel electrophoresis and detected by autoradiography. The total number of mitochondrially synthesized proteins analysed in this manner is estimated to be between 10 and 25 depending on the organism. Functional mitochondria contain at least 300-400 proteins, thus mitochondria are far from autonomous and a huge nuclear contribution in mitochondrial biogenesis and function is apparent. Indeed, not one of the major inner membrane respiratory complexes originates entirely from genes encoded in the mitochondria. They are therefore of dual origin, containing both nuclearly encoded proteins synthesized in the cytosol and transported into the mitochondria and other proteins which are synthesized endogenously before being assembled into a functional complex.

Complex I is the NADH:ubiquinone oxidoreductase (commonly referred to as 'the NADH dehydrogenase'), a large complex reported to consist of 16-18 subunits (Ise et al. 1985). In man, Drosophila and Neurospora crassa six subunits of
this complex are mitochondrially encoded and in plant mitochondria homology to
genes for three subunits has been found (compiled by Lonsdale 1988). No subunits
of complex II, the succinate:ubiquinone oxidoreductase, have yet been identified as
mitochondrial gene products in any organism. Complex III is the ubiquinol
cytochrome c oxidoreductase, comprising 8-10 subunits of which the
apocytocrome b protein, in all organisms investigated, appears to be the only
subunit of mitochondrial origin (Krieke et al. 1985). Complex IV, the cytochrome c
oxidoreductase (or cytochrome c) is a nine subunit assembly in yeast (Power et al.
1984) but may be larger in plants (Douce 1985). The subunits I, II and III of
cytochrome c are mitochondrially synthesized in both these and animal systems.

Complex V is the F_{1}F_{0}ATPase complex consisting of two portions, the F_{1} and
F_{0}, associated with the mitochondrial matrix and inner membrane respectively.
Both the F_{1} and F_{0} portions of the ATPase contain five polypeptide subunits.
Subunits 6 and 8 of the F_{0} portion are mitochondrially encoded in mammals and
fungi (Grivell 1989). In plants, sequences showing 60% sequence homology with
the subunit 6 genes from other organisms are present in mtDNA, but extensive
homology to subunit 8 appears to be lacking (Hiesel and Brennicke 1985). Genes
encoding subunit 9 of this complex have been found in plant mtDNAs although
they are nuclearly encoded in mammals and some fungi. In Neurospora crassa a
copy of the gene encoding subunit 9 is present in both the nucleus and the
mitochondria although the nuclear sequence appears to be the only functional one
(van den Boogaart et al. 1982). Plant mtDNAs additionally, and uniquely, encode
the α-subunit of the F_{1} portion of the ATPase (Braun and Levings 1985, Isaac et al.
1985a, Schuster and Brennicke 1986). The remaining components of the
mitochondrial membrane complexes, all the enzymes required for mtDNA
replication, RNA transcription, processing and protein translation plus those
involved in the catalysis of mitochondrial metabolic pathways are nuclearly
encoded gene products.

1.2.1 The Endosymbiont Theory

Many of the genes for mitochondrial proteins and tRNA molecules which
are nuclearly encoded may have been transferred during evolution from the
organellar genome. The mitochondrion is thought by many, to represent the
evolutionary relic of an ancestral prokaryotic organism which invaded some early
fermentative, nucleated cell. This so called 'endosymbiont theory' maintains that
invasion of the eukaryotic cell led to a symbiotic relationship of advantage to the
host due to more efficient energy production (Gray & Doolittle 1982). With time,
the prokaryote has degenerated, some of the functional coding regions of its
genome have been lost altogether (presumably because they were duplicated in the
nucleus), some have been transferred to the host nuclear genome and a limited
number have been retained in what today, is the mitochondrion. An alternative,
less widely accepted, hypothesis is that processes of intercellular subdivision have
given rise to the nuclear, mitochondrial and in photosynthetic cases, chloroplast
compartments of the cells of extant eukaryotes (Mahler and Raff 1975).

Similarities in the mechanism of protein synthesis found in mitochondria
and bacteria and sequence comparisons of the genes encoding the ribosomal RNAs
(rRNAs) in these systems have supported the implied prokaryotic origin which
forms the basis for the endosymbiotic hypothesis (Chao et al. 1984, Spencer et al.
1984). However, it is still obscure why during evolution such a small subset of
genes, and the ability to express these genes, have been retained within the
mitochondria. It may be due to the hydrophobic nature of the encoded proteins
which presents difficulties for transport across the membrane system although this
cannot apply to all the imported molecules. The mechanism by which the
postulated transfer of genetic information from the mitochondria to the nucleus
has occurred remains unclear and is discussed further in section 1.7.

1.2.2 Protein Import into Mitochondria

Polypeptides which are nuclearly encoded, cytosolically synthesized and
destined for the mitochondria show variable import pathways, in the few cases
which have been characterised. Our understanding of the mechanisms of protein
targeting and import comes mainly from studies performed with the fungi
*Saccharomyces cerevisiae* and *Neurospora crassa* (reviewed by Attardi and Schatz
1988). In these organisms, most nuclearly encoded mitochondrial proteins are synthesized with
amino terminal extensions that act as signals to direct the entire polypeptide to the
organelle. The exact requirements of a targeting sequence may vary, as the primary
sequence of such putative leader regions is not found to be highly conserved
between different proteins. However, those targeting sequences which have been
investigated share similarity in their predicted secondary structures. These are
alpha helical conformations that predict exposure of hydrophilic residues around
the helix structure which may be important in facilitating the transport process
(von Heijne 1986).

In yeast, the *in vitro* import of proteins into mitochondria is
post-translational and requires an electrostatic potential across the inner membrane
to initiate the process (Pfanner and Neupert 1986, Eliers et al. 1988). Energy in the
form of ATP or guanosine triphosphate (GTP) is also required to complete transfer
across the inner membrane, probably to maintain the precursor protein in an
unfolded state (Eliers et al. 1987). Unfolding is necessary to allow correct association
with membrane contact sites through which transport occurs in fungi and probably
higher plants (Schleyer and Neupert 1985). Proteins enter the matrix where the
targeting sequence is removed by proteolytic cleavage by a specific processing
activity. In yeast and *Neurospora* the matrix processing enzyme consists of a
catalytic component, the mitochondrial processing peptidase (MPP) and a
processing enhancer protein (PEP) which cooperate in the proteolysis (Yang et al. 1988). MPP and PEP are nuclearly encoded proteins and in Neurospora both are synthesised as precursors that are processed by their own mature counterparts during import into the mitochondria (Schneider et al. in press). Proteins such as cytochrome b$_2$ which are destined for the inner membrane space, enter the matrix and then traverse the inner membrane a second time. This is directed by a secondary signal sequence revealed by the initial proteolysis, which is in turn cleaved after the polypeptide has reached its final destination (Hartl et al. 1987).

In higher plants, the import pathways of cytoplasmically synthesized mitochondrial proteins is less well understood. Presequences at the front of two proteins from Nicotiana (tobacco), the beta ($\beta$) subunit of the F$_1$ portion of the ATPase and the mitochondrial superoxide dismutase, have been shown to contain targeting information for transport into tobacco or yeast mitochondria respectively (Boutry et al. 1987, Bowler et al. 1989). In maize, the adenine nucleotide translocator (ANT) is synthesized as a large precursor protein of around 38kDa, which when imported, in vitro, is processed to a mature polypeptide of 30kDa (C.J. Sarah, personal communication). The nature and position of targeting information within the maize ANT protein has yet to be elucidated. Recently it has demonstrated that a gene fusion containing the targeting presequence of the yeast mitochondrial tryptophanyl-tRNA-synthetase can direct a bacterial $\beta$-glucuronidase (GUS) reporter gene specifically into the mitochondria of transformed tobacco plants (Schmitz and Lonsdale 1989). This suggests that, in vivo, the mitochondrial protease of the plant system uses the same recognition site as the yeast MPP. It remains to be shown whether plant mitochondria contain separate processing activities for targeting proteins to the different organelle compartments analogous to the situation in chloroplasts where the import of certain proteins has been found to require stromal and/or thylakoid specific proteases (Kirwin et al. 1987).

1.3 Unique Features of Plant Mitochondria

Before making a comparison of the structure and expression of higher plant mitochondrial genomes with that of other organisms it is appropriate that some of the unique biochemical features of plant mitochondria, which have become apparent in the past few years, are mentioned. Plant mitochondria resemble their animal and fungal counterparts in the phospholipid and channel protein composition of their membranes and in essentially identical phosphorylating and anion translocating systems. However, distinct differences in the nature of the electron transport chain and in dehydrogenase activities exist in plants and merited a recent review by Douce and Neuburger (1989).

A cyanide (CN-) resistant respiratory activity, found in mitochondria of microorganisms and some fungi, but absent from animals, has been described in several higher plant species (Lloyd 1974, Elthon and McIntosh 1987). This CN-
resistant pathway is most simply described as a branch from ubiquinone in the conventional electron transport system which ends with a specific complex, distinct from the normal cytochrome oxidase, known as the alternative oxidase (Figure 1.1). Electron movement along this path bypasses two of the sites where H⁺ translocation is coupled to the synthesis of ATP. There appears to be no energy conservation associated with electron transport via the alternative oxidase and this has led to the postulation that the energy derived by this pathway is dissipated as heat (Moore and Bonner 1982). Indeed, thermogenesis is the only ascribed physiological role of respiration via the alternative oxidase. The spadices of Araceae species show very high rates of O₂ uptake, which is almost entirely cyanide insensitive indicating that the alternative oxidase pathway is engaged (Meeuse 1975). The associated rise in temperature releases volatile odiferous amines which attract insects for pollination. Although direct evidence is lacking, it has been suggested that the alternative oxidase acts as an overflow mechanism which is engaged when TCA cycle intermediates are required but the cytochrome pathway is constrained by high levels of ATP production (Day et al. 1983).

Complex I of the inner mitochondrial membrane of plants, and other organisms, shows rotenone inhibitor sensitive NADH dehydrogenase activity (see Figure 1.1). However, in plants rotenone insensitive NADH dehydrogenase activities are also present. A specific NADH dehydrogenase, situated on the outer surface of the inner membrane, has been found in all plant mitochondria investigated to date (Complex VII). This complex feeds electrons directly into the ubiquinone pool and then to complex III so bypassing the first site of H⁺ translocation. The combined activities of this dehydrogenase and the alternative oxidase means that a potential for electron translocation, which is completely uncoupled from ATP synthesis, exists in higher plants although exactly how and when this dehydrogenase activity interacts with the main respiratory chain is not understood. There is growing evidence that plant mitochondria preferentially oxidise endogenous NADH, produced in the matrix during the course of TCA cycle substrate oxidation, via a more complex process than occurs in animals. The clue, again provided by inhibitor studies, is that antimycin A only partially inhibits the process in plants, compared to the more or less complete inhibition it causes in animals. This, and similar observations, have led to the suggestion that at least two internal NADH dehydrogenases are present on the inner surface of the inner membrane in plants (Douce and Neuburger 1989).

Higher plant mitochondria have been found to readily oxidise malate in the presence of the oxaloacetate (OAA). As the product of malate oxidation, OAA inhibits the activity of the malate dehydrogenase enzyme in mammalian mitochondria if it is not rapidly removed from the mitochondria. In contrast, removal of OAA is not necessary in plant mitochondria because the matrix contains a specific NAD-dependant malic enzyme which catalyses the
decarboxylation of malate to yield pyruvate and CO₂ (Douce 1985). This reaction provides photosynthesis with CO₂ for the Calvin cycle and also represents a branch point for metabolism of malate in plant mitochondria. The NAD-malic enzyme allows continual turnover of the tricarboxylic acid (TCA) cycle when pyruvate is limiting, by ensuring the complete oxidation of stored reserves of malate via conversion to pyruvate and then acetyl-coenzyme A (acetyl-CoA). This conversion of C4 acids into acetyl-CoA, is an important aspect of the respiratory process in plant species which show C₄ and Crassulacean Acid Metabolism (CAM).

Many of the steps and functions of these complex biochemical processes which occur in plant mitochondria remain to be fully elucidated. Exactly how many, if any, of the proteins responsible for these distinctive characteristics of plant mitochondrial function are encoded within the mtDNA is unknown and awaits a complete description of the coding capacity of plant mtDNA. It is very likely that, as was found for components of the conventional respiratory chain, the majority of the proteins will be encoded in the nuclear genome.

The rest of this chapter describes our current understanding of the mitochondrial genetic system of higher plants in comparison to that of other organisms. Mutations associated with alterations in mitochondrial function which have been best characterised in fungi will be briefly outlined before an introductory discussion of the main subject of this thesis, namely the phenotypic trait in plants known as cytoplasmic male sterility.

1.4 Organellar Genomes - Organisation and Recombination

Studies in plant mitochondrial molecular genetics have lagged behind those in animal and fungal systems although it is already apparent that higher plant mitochondrial genomes differ from those of other organisms in a number of important aspects. Our understanding of the physical structure and information content of plant mtDNA has certainly increased over the last few years, however, only recently have studies progressed from the necessarily descriptive stages to investigations of the mechanisms by which the genetic system performs its crucial role.

1.4.1 The Structure and Organisation of Mitochondrial Genomes

The mtDNA of higher plants is much larger and more variable in size than that of animal and fungal species. The size complexity of plant mtDNAs investigated to date range from 208kb in *Brassica hirta* (white mustard) to 2400kb in *Cucumis melo* (muskmelon), (Ward *et al.* 1981, Palmer and Herbon 1987). This contrasts with the relatively invariant mammalian mitochondrial genome size of 15-17kb, those of protists and fungi ranging from 15-108kb and also the chloroplast genomes of plants and algae which, with few exceptions, are between 120-210kb (Newton 1988, Palmer 1985). In animals, the mitochondrial genome is organised as
a single circular molecule and, in those genomes which have been completely sequenced, it is apparent that a very compact arrangement of genes has been highly conserved during evolution (Attardi 1985). The genes of mammalian mtDNA do not contain intervening sequences (introns) and in most cases very few, if any, nucleotides separate adjacent coding regions, resulting in an extremely economical organisation of the coding information (Anderson et al. 1981). In protozoa, genomes are organised either as circular chromosomes, such as the 18.5kb maxicircle of *Trypanosoma* (Simpson 1987) or as linear double-stranded molecules with terminal repeats like the 31.5kb genome in *Tetrahymena* (Morin and Cech 1986).

The mitochondrial genomes of fungi are more variable. Strains of the fission yeast *Saccharomyces pombe* have genomes of 17-19kb whereas those of bakers yeast, *S.cerivisiae*, are 75-85kb, although both are organised as circular molecules. The size variation is partly due to noncoding adenine-thymine (AT) rich intergenic regions and 'optional' introns such as the b13 intron of the apocytochrome b gene which produces long or short forms of the genome (Grivell 1989). The variable presence of introns also accounts for differences in the mtDNA complexity of strains of filamentous fungi such as *Neurospora crassa*, which contain mitochondrial genomes varying from 60-73kb (Burke and RajBhandary, 1982, Wallace 1982). Size differences are greater when more distantly related species are compared, such as *Aspergillus nidulans*, a deuteromycete fungus which has a mitochondrial genome of 33kb (Brown et al. 1985), in comparison with the 94kb mitochondrial genome of the ascomycete *Podospora anserina* (Cummings et al. 1979).

The unique ability of yeast to survive anaerobically with defective mitochondria has been exploited in analyses of mutations which affect mitochondria function and to develop our understanding of the mitochondrial coding capacity in this organism. The so called 'petite' mutants of yeast lack respiratory function and grow as small colonies when a fermentable carbon source is limiting. In petites, the mitochondrial genome consists of excised segments of the parental wild-type genome, which are tandemly amplified to become a repeat unit within the petite genome (Gaillard et al. 1980). The excision sequences involved in petite genome formation are localised in 'AT spacer' or 'GC cluster' noncoding regions that comprise over 50% of the yeast mtDNA. Perfect direct repeats within these regions are involved in the excision process and the abundance of such sequences within the mitochondrial genome of wildtype yeast accounts for the high frequency of the spontaneous petite mutation (de Zamaroczy et al. 1983).

Defective mitochondrial genomes derived by deletion and amplification by recombination across short repeated sequences also occur in the 'poky' and 'stopper' mutants of *Neurospora* (Betrand et al. 1980) and the ragged mutants of
Aspergillus amstelodami (Lazarus and Kunzel 1981). An amplified segment from an Aspergillus ragged mutant has been found to replicate within a yeast host cell and the putative origin of replication responsible for amplification has been delimited (Beri et al. 1988). Specific mtDNA rearrangements characterise the senescence phase in the life cycle in Podospora anserina (Koll et al. 1985). In Podospora mutants which do not undergo senescence, insertions of short stretches of A and T residues occur in the same region of mtDNA which is involved in the senescence process and have led to the postulation that recombinational 'hotspots' may determine the position of some mitochondrial rearrangements (Koll et al. 1987).

In common with the mammalian mitochondrial genomes, yeast mtDNA contains genes encoding the large and small subunit rRNAs, a set of transfer RNAs (tRNAs) and several polypeptide components of the respiratory enzyme complexes. Additionally, there are genes having no identifiable counterparts in mammalian mtDNAs. These include enzymes associated with RNA processing and maturation and a protein component of the small ribosomal subunit, called var1 (Grivell 1989, Butow et al. 1985). In contrast to the six subunits of complex I NADH dehydrogenase encoded in mammalian mtDNA (Chomyn et al. 1985) no homologues in yeast mtDNA have been found. A recent compilation of all the available sequence for S.cerevisiae, covers over 92% of the entire genome and renders it unlikely that much more in the way of coding sequences will be found (Grivell, 1989).

1.4.2 Higher Plant Mitochondrial Genomes

The number of mitochondria per cell varies considerably in plants and other organisms and is assumed to reflect the respiratory requirement of the tissue in which they occur (Bendich and Gauriloff 1984). Only a few reported studies relate the actual mtDNA content to the number of mitochondria in a cell. In one of the earliest studies, Suyama and Bonner (1966) estimated $5 \times 10^{10}$ µg DNA to be present per mitochondrion in several different species (turnip, sweet potato, onion and mung bean). The sizes of mitochondrial genomes in a number of plants have now been accurately estimated by physical mapping studies. Among the Brassica species mitochondrial genome size is relatively conserved at around 200-250bp (Palmer and Herbon 1987). Assuming a size of of 220kb for the turnip (Brassica rapa) genome the calculated DNA concentration per single mitochondrion, in this species, equates to two complete mitochondrial genomes. However, in some Cucurbitaceae, the presence of thousands of mitochondria per cell, but mtDNA equivalent to only 110 genomes, has led to conclusion that less than one complete genome is present per mitochondrion in these species (Ward et al. 1981). This lends support to the proposal by Lonsdale et al. (1988) that the mitochondria of plants, like yeast, form a continuous syncytium and that mtDNA within individual
cells behaves as population of interrecombining molecules rather than a collection of discrete entities within single organelles. This is not implausible given the 'fluidity' in the physical organisation of mitochondria described earlier. Cell fusion studies in Nicotiana (tobacco) provide direct evidence that mitochondrial fusion is an active process in higher plant cells which would be expected to influence how the mitochondrial genome is distributed, maintained and inherited (Lonsdale 1987).

An analysis of mtDNA of several species from the Cucurbitaceae has revealed a seven fold range in genome sizes within this single higher plant family from 330kb in watermelon to 2400kb in muskmelon (Ward et al 1981). In the cucurbits examined, the mitochondrial genomes contain less than 10% repeated DNA which is therefore not a major contributor to the size increase. Attempts to explain the large size have ranged from the possibility that plant mtDNA contains many more genes and/or regulatory sequences than occur in fungal and animal mtDNAs, to the suggestion that a large proportion of the mtDNA may have no function at all, existing merely as a consequence of the self-perpetuation of DNA, so-called 'selfish DNA' (Cavalier-Smith 1987). Both of these explanations have, to a large extent been discounted and, as characterisation of plant mtDNA continues, unique structural features which affect the size and organisation of these genomes are becoming apparent (Levings and Brown 1989). In particular, there has been the finding that plant mitochondrial genomes show a relatively high frequency of inter-molecular and intra-molecular recombination which lead to extensive rearrangement of a basic genomic structure.

Early analysis of plant mitochondrial genomes by electron microscopy indicated that the mtDNA is composed of an array of differently sized circular and linear molecules (Synenki et al. 1978). The lengths and proportions of these mtDNA molecules vary with the species and tissue source examined. Circular configurations have been found to represent a greater proportion of the total mtDNA isolated from plant cell tissue culture rather than intact plant tissue and in Oenothera and tobacco different size classes of circular molecules found are due, in part, to multimeric association of smaller circles (Brennicke and Blanz 1982, Dale et al. 1983).

An explanation of the observed heterogeneity has been provided by detailed physical mapping studies of the mitochondrial genomes from several higher plant species (Lonsdale et al. 1984, Quetier et al. 1985, Palmer and Shields 1984). Plant mitochondrial genomes can be depicted as circular molecules which contain the entire sequence complexity of the mtDNA. However, due to the presence of repeated sequences within the main genomes, subgenomic molecules can arise by recombination across the repeats and lead to a more complex organisation. The mitochondrial genome in Brassica campestris consists of a main or 'master' circle of 218kb which includes a 2kb direct repeat (Palmer and Shields 1984). Two smaller
circular molecules of 153kb and 83kb are also present at lower abundance relative to the master circle (i.e. lower stoichiometry). These smaller circles can be accounted for by postulating recombination across the direct repeat which divides the master circle into two subgenomic circles, the sizes of which are determined by the distance between the copies of the repeat on the parental molecule (see Figure 1.2a). The organisation is described as 'tripartite' since three components of the genome conformation are present in recombinational equilibrium. Such a tripartite organisation has also been described for the 305kb mitochondrial genome of sunflower in which recombination across a 12kb direct repeat can generate subgenomic circles of 241kb and 64kb as shown in Figure 1.2b (Siculella and Palmer 1988).

More complex organisations have been reported for the 570kb mitochondrial genome of N type maize (Lonsdale et al. 1984), the 450kb genome of wheat (Quetier et al. 1985) and the 386kb genome of the sugar beet (Lonsdale et al. 1988). Many subgenomic molecules are generated in these species as a consequence of their larger size and greater number of repeats. Most repeated sequences in plant mtDNA are present in direct orientation on the main genome although indirect repeats are found in maize, sugar beet and Petunia mtDNA (Lonsdale et al. 1984, 1988, Folkerts and Hanson 1989). In maize, five direct repeats and one indirect repeat have been mapped on the 570kb master circle (see Figure 1.2c). Recombination across the 12kb direct repeat in the master circle predicts the formation of a 67kb subgenomic circle. A circular molecule of this size is the predominant conformation observed by electron microscopy while the subgenomic circle of 503kb generated by the same event, and the 570kb master circle, most probably fragment during isolation of the mtDNA. Subsequent recombinations between the products of this, and other recombination events give rise to a highly heterogeneous or multipartite organisation for the maize genome (1.2c).

In Petunia three copies of a recombinationally active repeat have been located on a map of the 443kb mitochondrial genome (Folkerts and Hanson 1989). Two copies of a 6.6kb sequence are in a direct orientation and a third 3.5kb truncated version is inverted relative to the other two. The presence of these repeats predicts a multipartite molecular organisation for Petunia mtDNA consisting of four alternative master circles and three subgenomic circles. The smallest plant mitochondrial genome yet described is the 208kb of Brassica hirta (white mustard) (Palmer and Herbon 1987). This genome is devoid of active recombinational repeats and exists as a single circular molecule, proving that a high degree of organisational complexity is not the unexceptioned rule in plant mtDNA.
Physical mapping studies have been used to identify linkage groups within mtDNA from several plant species and to describe the organisation of their mitochondrial genomes. In *Brassica campestris* (turnip) (a) and *Helianthus annuus* (sunflower) (b), the genomes contain single repeats of 2kb and 12kb respectively. The direct orientation of the repeats on the ‘master circles’ predicts that recombination across the repeats will produce two smaller subgenomic circles as part of the *tripartite* structure of these genomes (Palmer and Shields 1984, Siculella and Palmer 1988). This type of organisation is common among *Brassica* species, which are generally of 200-250kb in size (Palmer and Herbon 1986). The mitochondrial genome of N (fertile) type maize is 570kb (c) and contains 6 repeats of at least 1kb in size. This leads to the formation of numerous subgenomic circles as a result of multiple recombination events, only a subset of which is represented in this figure. Such *multipartite* organisations are also found in wheat (Quetier *et al.* 1985), *Petunia* (Folkerts and Hanson 1989) and sugar beet (Lonsdale *et al.* 1988).
Figure 1.2 Variation in Plant Mitochondrial Genome Organisation

a. Tripartite organisation of the mitochondrial genome of *Brassica campestris*

- 2kb direct repeat
- 83kb
- 135kb
- 218kb master circle
- recombination intermediate
- subgenomic circles

b. Tripartite organisation of the mitochondrial genome of *Helianthus annuus*

- 12kb direct repeat
- 305kb
- 64kb
- 241kb

C. Multipartite organisation of the mitochondrial genome of *Zea mays*

- 1kb, 2kb, 3kb, 10kb, 12kb direct repeats
- 14kb indirect repeat
- 570kb
- 488kb
- 250kb
- 253kb
- 456kb
- 503kb
Despite the apparent capacity for generating organisational diversity, plant mitochondrial genome structure is remarkably stable and shows only low levels of intraspecific variability (Palmer and Herbon 1988). Transmission of plant mtDNA from generation to generation, via the maternal cytoplasm, is highly conservative as is apparent from the identical restriction endonuclease patterns for mtDNA from progenitor and progeny plants. Tissue culture manipulations, however, can promote extensive rearrangement in plant mtDNAs and have been used as systems to study the nature and frequency of mtDNA recombination in cultured cells from a number of species (Hartmann et al. 1989, Shirazadegan et al. 1989). In *Brassica campestris* a number of mtDNA rearrangements which derive from tissue culture manipulations have been found occur in the same region of the genome as three 'evolutionary' inversions which distinguish the genomes of *B. campestris* and *B.oleracea* (Shirazadegan et al. 1989, Palmer and Herbon 1988). The suggestion from this observation is that certain regions of this genome, and possibly plant mtDNAs in general, may have a greater 'tendency' to promote restructuring of the genome by intramolecular recombination.

Subgenomic molecules which, in part, account for the observed complexity of plant mitochondrial genomes must be replicated coordinately with the master circle at rates which maintain their copy numbers for passage to the progeny. From studies of whole plant and culture derived material, it is now apparent that there is yet another aspect to the complexity. Some mtDNA molecules, apparently representing further arrangements of the genome, persist from generation to generation at very low levels relative to the more apparent forms of the genome (Small et al. 1987). In order to be maintained these molecules which are present at low stoichiometry must replicate independently of the main genomic and subgenomic molecules. In maize, cytoplasms previously characterised by the presence of specific mitochondrial genome organisations, have been found to contain alternative organisations at very low levels and this has led to postulation that amplification of such pre-existing non-abundant molecules (which have been termed 'sublimons') may be the method by which organisational alterations occur during genome evolution (Small et al. 1989).

It is of note that, since the original reported finding of sublimons in maize mitochondria isolated from seedling tissue, similar observations of substoichiometric organisations have been reported in tissue culture material from a variety of plant species. Ozias-Akin et al. (1988) have found that apparently novel mtDNA restriction endonuclease fragments in somatic hybrids produced by protoplast fusions between two different species *Panicum* and *Pennisetum* are actually present at low copy number in one of the parents. Similarly, in *Brassica* tissue culture material, at least some of the mtDNA molecules which appear to become amplified following long-term *in vitro* culture are actually present in the parental plant, but at much lower (1-5%) abundance (Shirazadegan et al. 1989).
implies that various alternative products of recombination events may be retained in plant mitochondrial genomes for many generations at low levels. Under certain circumstances, e.g. in vitro culture, fusion with another cytoplasm or introduction of the cytoplasm into a new nuclear background, some form of amplification and assortment of subliminal molecules already present in the plant tissue may occur.

1.4.3 Mechanisms of Genome Recombination

The proposed recombinational 'flexibility' of the plant mitochondrial genomes is used as the explanation of the observations from physical mapping and electron microscopy. However, there has been as yet no direct demonstration, in vivo or in vitro, of recombination in plant mitochondrial genomes. In fungi, three recombination mechanisms which have been proposed to account for mitochondrial genome rearrangements are general recombination, site-specific recombination and transposition.

General recombination occurs when there is genetic exchange between two homologous DNA strands promoted by a general recombinase activity, whereas site-specific recombination is mediated by a recombinase protein which recognises specific sequence elements at or near the point of recombination (see Figure 1.3a). These elements may be a particular primary sequence, or a conserved secondary structure such as a hairpin loop configuration derived by DNA folding. Transposition is a general term relating to the movement of genetic material by replication of the transposable sequence (transposon) at its original position and integration of the copied sequence elsewhere. A further mechanism of rearrangement is reverse transcription/ligation. This involves an RNA intermediate which acts as a template for the synthesis of a complementary DNA (cDNA) molecule which subsequently becomes ligated back into the genome. The possibility that reverse transcription occurs in plant mitochondria and is involved in genome rearrangement is discussed in a later section (1.7).

Several of the repeats found in higher plant mtDNA have been sequenced (Houchins et al. 1986, Hiesel et al. 1987) and only a few contain regions of homology to those known to promote site-specific recombination in other organisms. A decanucleotide sequence (5'GGAAGCAGCC3') found within the 26S rRNA gene of Oenothera is the site for a recombination event involved in the formation of a 7.5kb subgenomic circle in this genome. This sequence shows some similarity to the omega insertion site sequence found in the 21S rRNA gene of yeast mtDNA (Dujon 1980) although such comparative analysis is speculative in the absence of a functional test of the plant sequences. Generally, there is a lack of sequence homology between different repeats within individual genomes, and between repeats found in different species (Lonsdale et al. 1988). Hybridisation studies in Brassica and Raphanus demonstrate that repeated DNA sequences in one species
exist as single copy in others (Palmer and Herbon 1986). This suggests that specific primary sequences are not required for plant mtDNA recombination and that the presence of a duplication is in itself sufficient. This in turn implies the operation of a general recombination mechanism and it is true that much of the observed complexities of plant mitochondrial genomes can be explained by invoking reversible homologous recombination. However, the fact that some repeats in plant mitochondria appear to be recombinationally inactive cannot be reconciled with such a general mechanism. For example, none of the predicted products of recombination across the 10kb repeat in maize have been detected among the subgenomic circles (Lonsdale et al. 1988) and in Petunia, repeats have been found which are apparently neither substrates for, nor products of, detectable recombination events (Folkerts and Hanson 1989). It is also becoming apparent that certain short repeats, present in multiple genomic environments within the mtDNA of different plant species, do share homology (Bland et al. 1987, Hiesel et al. 1987, Folkerts and Hanson 1989 and discussed in chapter 4). The implication is that the recombination processes in plant mitochondrial genomes may entail a degree of sequence specificity which is not yet understood.

As stated, homologous recombination across repeated sequences does provide a suitable explanation of the presence of reciprocal recombination products among complex plant mitochondrial genomes. In plant mtDNA each copy of a repeated sequence is flanked by a pair of single copy regions which can be used to distinguish the parental and recombined configurations of the mtDNA following recombination (see Figure 1.3). If a sequence is directly repeated once, four combinations of sequences flanking the repeat would produce four genomic environments as a result of recombination within the repeat (a). Within circular genomes, recombination across direct repeats predict the formation of subgenomic circles by 'looping out' (as shown in Figure 1.2) whereas recombination across indirect repeats leads to inversion of the region between them by 'flip-flopping' (Figure 1.3b).

1.4.4 Chloroplast Genomes

The organisation of the chloroplast genome is generally conserved in both size (120-200kb) and gene order among higher plants and is characterised by the presence of an inverted repeat (ca. 21-28kb) which divides the genome into large and small single-copy regions (Palmer 1985). The relatively small range of chloroplast DNA (cpDNA) complexities is thought to reflect strong constraints on the size of chloroplast genomes during evolution although the nature of such constraints remains unclear (Palmer et al. 1987). Intramolecular recombination between the chloroplast inverted repeats has been shown to occur in a number of diverse species and results in two genetically identical but physically distinct isomeric forms of the genome (see Figure 1.3b). Observations of substoichiometric
cpDNA organisations, similar to those reported for mitochondrial genomes have also been made (Moon et al. 1987).

The complete physical map of the sunflower cpDNA has been reported to be 152bp in size, consisting of an large (86kb) single-copy region and a small (20kb) single-copy region (Heyraud et al. 1987). The physical map of of sunflower cpDNA is almost colinear with that reported for the tobacco chloroplast genome (Shinosaki et al. 1986). However, within the large single-copy region of sunflower cpDNA, a region which encodes the genes for several tRNA genes differs in organisation when compared to cpDNA from tobacco and several other plant species. The altered arrangement in the sunflower cpDNA can best be explained by an inversion of a region, spanning approximately 23.5kb which extends from the sites corresponding to two tRNA genes on the tobacco genome. As a consequence of this rearrangement, the atpA gene, which is included within the inverted region, has a reversed orientation within sunflower cpDNA relative to tobacco cpDNA. Sunflower is a member of the Compositae family and the cpDNA from three other members, safflower, Lactuca sativa (lettuce) and Barnadesia caryophylla have been found to contain similar inversions involving the atpA gene (Ma and Smith 1985, Jansen and Palmer 1987).

The chloroplast genome of geranium at 217kb is unusually large and shows a unique combination of structural alterations relative to other land plants (Palmer et al. 1987). Much of the extra size is due to the 76kb inverted repeat which is three times larger than the 'normal' chloroplast inverted repeat. Many regions which are single copy in other cpDNAs are therefore duplicated in the geranium cpDNA, including the sequences for ten protein coding genes. No fewer than six inversions, occurring in both the large single-copy and inverted repeat regions are are necessary to align the geranium chloroplast genome with the organisation more generally found in plant cpDNAs. Homologous recombination between the unusually large number of dispersed repeats found within the geranium cpDNA is thought to be a major contributing factor to the degree of inversion found within this genome (Palmer et al. 1987). Although such recombination can, to some extent, explain the apparent extensive restructuring of the chloroplast genome in geranium, the preponderance of dispersed repeated sequences is unusual. As stated earlier, recombinationally active inverted repeats occur within the mtDNA from Petunia and sugar beet where they are also found to contribute to the overall genomic complexity (Folkerts and Hanson 1989, Lonsdale et al. 1988).
Figure 1.3 Recombination across repeated sequences

a. General or Site-specific Recombination
Recombination across a repeat, flanked by four unique single copy sequences in an original configuration, results in linkage of alternative combinations of the flanking regions of the repeat in the derived configuration.

b. Recombination across Indirect Repeats
Recombination across an indirect repeat within a circular genome leads to inversion of the sequence between the repeats by 'flip-flop' mechanism.
1.5 The Coding Capacity of Plant Mitochondrial Genomes

Plant mitochondrial genomes contain the same basic set of genes identified in the mtDNA of animals and fungi (Newton 1988). Additional genes, not found to be mitochondrially encoded in other organisms, have also been found in plant mitochondria.

1.5.1 Mitochondrial Ribosomal RNA genes

Mitochondrial ribosomes in higher plants differ from those in animals and fungi. They contain a 26S large subunit ribosomal RNA (rRNA) and an 18S small subunit rRNA which are both larger than their counterparts in mitochondria of fungi and animals and in the chloroplast ribosomes (Dale et al. 1984, Chao et al. 1984). Plant mitochondria also uniquely contain a small 5S rRNA (Leaver and Harney 1976). In all plant species examined to date, the genes for the 26S rRNA (rrn26), 18S rRNA (rrn18), and the 5S rRNA (rrn5) are mitochondrially encoded. In addition, the rrn18 and rrn5 genes are closely linked and are found 108bp apart in maize (Chao et al. 1983) and 582bp apart in Oenothera (Brennicke et al. 1985). However, the rrn26 gene is located several kilobases from this linkage group (Stern et al. 1982). A separate location for the large and small subunit genes is also found in yeast and Aspergillus whereas in mammalian mtDNA these genes are closely linked and cotranscribed (Grivell 1989, Attardi et al. 1985).

The level of nucleotide sequence divergence found between plant mt rRNA genes is very low. Maize and wheat rrn18 genes are 96% homologous (Chao et al. 1984, Spencer et al. 1984). Sequence analysis of the mitochondrial 18S gene in soybean, a dicotyledonous species, has shown that it is also remarkably similar (85%) to those found in the two monocots and suggests that plant mtDNA shows a slow rate of nucleotide divergence in comparison to mtDNA of other organisms (Grabau 1985 and section 1.5.3). Also, both the primary sequence and proposed secondary structure of the plant 26S and 18rRNAs show considerable similarity to their counterparts in E.coli and chloroplasts, and have been used as supporting evidence for the eubacterial endosymbiotic origin of the plant mitochondria (Gray et al. 1984).

1.5.2 Mitochondrial Transfer RNA genes

The coding location of mitochondrial transfer RNA (tRNA) genes is variable. In fungi and animals all the mitochondrial tRNAs are encoded within the organelle, however the mitochondrial genomes of Chlamydomonas and Tetrahymena species do not contain a full set of tRNA genes and some are encoded in the nucleus and must be transported into the mitochondria (Suyama 1986, Gray and Boer 1988). Sequences for tRNA genes tend to 'punctuate' regions which encode protein coding genes in the mitochondrial genomes of mammals and some fungi, but a more clustered organisation of tRNA genes is found in the
mitochondrial genome of *S. cerevisiae* (Clayton 1984, Grivell 1989).

Genes for sixteen different tRNAs specifying eleven amino acids have been identified in plant mitochondria (reviewed by Lonsdale 1988). These tRNA genes are dispersed around the mitochondrial genomes and show little conservation in the order in which they occur in different species. As in yeast mitochondria, higher plant mitochondria contain separate initiation and elongation specific methionine tRNAs (Gray and Spencer 1983, Parks *et al.* 1984, Marechal *et al.* 1986). This contrasts with mammalian mitochondria, in which only one methionine tRNA is found (Attardi *et al.* 1985). Genes which encode separate isoaccepting tRNAs for the amino acid residue serine have been also been identified in the mitochondrial genome of wheat (Joyce *et al.* 1988).

Sequence analyses have supported claims that some mitochondrial tRNAs have been derived from the chloroplast. In maize, the mitochondrial cysteine tRNA (*trnC*) gene shows strong sequence homology (97%) to the corresponding gene which occurs in the chloroplast. Homology between the two organellar genomes extends beyond the coding region of the *trnC* gene, suggesting that this sequence in maize mtDNA may represent a chloroplast insertion into the mitochondrial genome (Wintz *et al.* 1988). Similarly, in wheat, the tryptophan tRNA (*trnW*) genes in the mitochondrial and chloroplast genomes are highly homologous and differ by only three nucleotides (Marechal *et al.* 1987). In addition, the 5' region of the mitochondrial *trnW* contains part of the proline tRNA (*trnP*) in a linkage identical to that found in the chloroplast genome, and which supports a plastid origin for this sequence. In maize, the *trnW-trnP* linkage is also found as a cpDNA insertion, and is present on a 2.3kb mitochondrial plasmid which characterises the cytoplasms of this species (Leon *et al.* 1989 and discussed in section 1.6). These 'transposed' tRNA sequences do appear to be transcribed within mitochondria unlike other cpDNA homologies which characterise the mitochondrial genomes of higher plants (Fejes *et al.* 1988, Stern and Palmer 1984).

In mungbean mitochondria four tRNAs which specify leucine are nuclear gene products which are specifically imported into the mitochondrial compartment (Green *et al.* 1987). The nucleotide sequences of these tRNAs, destined for the mitochondria, are identical to that of the leucine tRNA which functions in cytosolic protein synthesis but are imported as a result of a post-transcriptional modification involving methylation of particular G residue (Marechal-Drouard *et al.* 1988). Although it is possible that the number of tRNA genes in plant mtDNA may exceed the 22-25 encoded by mammalian and yeast mtDNAs (Attardi *et al.* 1985, Tzagaloff and Myers 1986), it seems likely that additional tRNAs, unaccounted for in plant mitochondrial genomes, will also be found to be imported from the cytosol.
1.5.3 Mitochondrial Protein Coding Genes

Mitochondria isolated from plants synthesize of the order of 20-25 polypeptides *in vitro* which is higher than is found in either animals (13) or fungi (ca.15) (Leaver et al., 1988). However, this apparent increase in protein coding capacity cannot fully account for the large and variable size of plant mitochondrial genomes described earlier. In higher plants, mitochondrial genes are dispersed around the genome, generally in an order which varies even between closely related species, although some indications of conserved gene linkage groups are emerging (Gualberto et al. 1988, Wissinger et al. 1988).

Open reading frames (ORFs) for thirteen different polypeptides have been identified in plant mitochondria by heterologous probe hybridisation and sequence analysis. Table 1.1 compiles published sequences for plant mitochondrial protein genes, and compares the coding location of these sequences in animals and fungi. Functions have been assigned to all 13 of these ORFs in plant mtDNAs, mainly by comparison of the predicted amino acid sequences with those of known proteins from other organisms. These include *cob, coxI, coxII, coxIII, atp6, nad1, nad3* and *nad5* which encode apocytochrome b of the bc1 complex, subunits 1, 2 and 3 of the cytochrome c oxidase complex, subunit 6 of the F1-F0 ATPase and subunits 1, 3 and 5 of the NADH dehydrogenase respectively, all of which are encoded in the mitochondria of plants, animal and fungi. However, as stated earlier, plant mitochondrial genomes also contain genes for the subunits 9 and α of the F1-F0 ATPase (*atp9* and *atpA*) which are nuclear gene products in mammals and some fungi (van den Boogaart et al. 1982, Brown et al. 1985). In cucumber, an 8kDa mitochondrialy synthesised protein has been identified as subunit 9 by its capacity to bind the compound dicyclohexylcarbodiimide (DCCD) (Hack and Leaver 1984). In maize and pea, immunoprecipitation of *in organello* translation products with an antibody raised against the yeast α subunit was used to confirm the mitochondria as the site of synthesis of this protein in these species (Hack and Leaver 1983, Boutry et al. 1983). Immunodetection with antibodies raised against the corresponding yeast proteins has also been used to identify the mitochondrial COXI in sorghum (Bailey-Serres et al. 1986) and ATP9, COXI and COXII polypeptides in maize (Liddell and Leaver, unpublished) among the products of *in organello* translation.

Our understanding of the structure of genes encoded within higher plant mitochondrial genomes has been greatly improved by sequence analysis and mapping studies to characterise these coding sequences. In general, genes are scattered throughout the genomes, and apparently independently transcribed. Many plant mitochondrial genes show uninterrupted coding regions although more complex organisations have been reported for the *coxII* and the *nad* subunit gene in several species. The *coxII* genes from maize (Fox and Leaver 1981), rice (Kao et al. 1984) and wheat (Bonen et al. 1984) all contain a single intervening (intron)
sequence which is missing in the corresponding genes in pea and *Oenothera* (Moon *et al.* 1985, Hiesel and Brennicke 1983). Extended versions of this intervening sequence have recently been reported for the *coxII* genes which occur in the mtDNA from sunflower and *Petunia* (Gallerani *et al.* 1988, Hanson *et al.* 1988a). The *coxII* intron contains conserved sequences typical of group II introns which are involved in the correct alignment of exons during the splicing reaction.

Within mtDNA from several plant species, an ORF showing homology to the central and carboxy terminus of the *nad1* sequence from fungi has been found but in all cases homology corresponding to the amino terminus of the protein is missing (see Table 1.1). Gene products corresponding to the *nad1* sequence have not been reported and it is unclear whether this gene is functional within plant mitochondria (Levings and Brown 1989). Markaroff and Palmer (1987) have shown that homology to *nad1* is present in the mitochondrial genome of *B.campestris* but is untranscribed and it has been suggested that the functional copy of this gene may reside in the nucleus in this, and other, species. In broadbean, an intron within the *nad1* homology contains an unusually large ORF which shares some homology with the sequence of an RNA maturase enzyme encoded by the first intron of the *coxII* gene in *S. cerevisiae* (Wahleithner and Wolstenholme 1988b, Grivell 1989). The *nad1* intron ORF sequence also shows some features associated with reverse transcriptase enzymes found in retroviruses although whether the ORF is expressed and has any functional significance in plant mitochondrial genome expression is unknown.

Sequence analysis of the *nad5* gene in *Oenothera* and wheat shows it also contains an intron (Wissinger *et al.* 1988, Bonen *et al.* 1988). Three linkage groups containing *nad5* homology occur in *Oenothera* mtDNA. One is an entire copy of the gene, linked with the genes for the 18S and 5S rRNA. The two additional regions of homology have apparently been generated by separate recombination events, in one case involving the first intron of *nad5* (Wissinger *et al.* 1988). Homology to *nad4* has recently been reported in mtDNA from wheat and *Petunia* although whether the genes show complex structures like the other mitochondrial encoded subunits of complex I remain to be shown (Gualberto *et al.* 1989, Folkerts and Hanson 1989).

In some fungi, ribosome-associated proteins similar to those found in *E.coli* are encoded by the mitochondrial genome. In *S.cerevisiae* and *N.crassa* the small subunit ribosomal proteins, var1 (Butow *et al.* 1985) and S5 (Lambowitz *et al.* 1979) respectively, are mitochondrial gene products but analogous genes are not encoded by animal mtDNA (Schieber and O'Brien 1985). In several plant species, mitochondrial genes which show varying degrees of homology to the *E.coli* ribosomal proteins have been identified. Ribosomal protein 13 (*rps13*) homology occurs in *Oenothera*, tobacco, maize and although also present in wheat, is apparently untranscribed during early stages of development (Schuster and
Homology to *rps13* is not detectable in the mitochondrial genomes of pea or broadbean and an active *rps13* gene may be nuclearly encoded in these and other species. In both tobacco and maize, *rps13* is linked with the gene for *nad1* (Bland *et al.* 1986). Another ribosomal protein gene, *rps12*, also appears to be part of a conserved arrangement as it has been found to be closely linked with *nad3* in maize, wheat and *Petunia* (Gualberto *et al.* 1988, Folkerts and Hanson 1989). These preliminary observations suggest that an element of 'regional conservation' in some plant mitochondrial genomes is becoming apparent as the search for more genes continues.

Exactly how many more functional mitochondrial genes exist in higher plant mtDNAs, and what contribution they make to genome size and variation, remains to be established. It is apparent that the amount of repeated sequence accounts for variable proportions of the total genome complexity between plant species (Palmer and Herbon 1987, 1988). Hybridisation studies and sequence comparisons across species boundaries indicate that a high degree of homology exists between plant mtDNA coding regions in different genomes but that there is apparently little conservation between non-coding regions (Clarke-Walker 1985). The variation in plant mtDNA intergenic regions may reflect extensive rearrangement by recombination so that they no longer appear homologous between species. Such a 'scrambling' of short sections of mtDNA throughout the genome would result in variable levels of organisational complexity without significantly altering the base composition of the primary sequences involved.

Higher plant mtDNAs do show remarkably uniform GC contents of ca. 47% and base substitution rates calculated to be one hundred times lower than those of vertebrate mtDNA. This has led some to postulate that the mitochondrial genomes of plants are evolving more slowly, at the nucleotide level, than those of other organisms (Palmer and Herbon 1988). If the slow base mutation rate of plant mtDNA constitutes an evolutionary disadvantage, it may be that the capacity to reorganise genomic structure through recombination, which will periodically create expressed genetic diversity, has been selected during plant mitochondrial genome evolution. Palmer (1987) summed up our current understanding very well in his statement that "Plant mitochondrial genomes change very slowly in base sequence but very rapidly in structure and arrangement. The mitochondrial genome can be viewed as a collection of unchanging sequences whose relative arrangement is very fluid."
<table>
<thead>
<tr>
<th>Complex/Subunit</th>
<th>Gene</th>
<th>Animals</th>
<th>Fungi</th>
<th>Plants</th>
<th>Species/Reference†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome oxidase</td>
<td>Subunit I*</td>
<td>coxI</td>
<td>-</td>
<td>+</td>
<td>+ maize,<em>Oenothera</em> sorghum, wheat (1-4)</td>
</tr>
<tr>
<td>Subunit II *</td>
<td>coxII</td>
<td>-</td>
<td>+</td>
<td>+ maize, maize:cms-C, wheat, rice, pea, soybean, <em>Oenothera</em> (5-11)</td>
<td></td>
</tr>
<tr>
<td>Subunit III</td>
<td>coxIII</td>
<td>-</td>
<td>+</td>
<td>+ <em>Oenothera</em> (12)</td>
<td></td>
</tr>
<tr>
<td>Cytochrome bc₁</td>
<td>apocytochrome b</td>
<td>cob</td>
<td>-</td>
<td>+</td>
<td>+ maize, <em>Oenothera</em>, wheat broadbean (13-16)</td>
</tr>
<tr>
<td>F₁-F₀ ATPase</td>
<td>Subunit α (F1) *</td>
<td>atpA</td>
<td>-</td>
<td>-</td>
<td>+ maize, maize:cms-T <em>Oenothera</em>, tobacco, (17-20)</td>
</tr>
<tr>
<td>Subunit 6 (F0)</td>
<td>atp6</td>
<td>+</td>
<td>-</td>
<td>+ maize:cms-T, maize-C soybean, <em>Oenothera</em> (21-24)</td>
<td></td>
</tr>
<tr>
<td>Subunit 8 (F0)</td>
<td>atp8</td>
<td>+</td>
<td>+</td>
<td>? <em>Oenothera</em> (25)</td>
<td></td>
</tr>
<tr>
<td>Subunit 9 (F0) *</td>
<td>atp9</td>
<td>-</td>
<td>+/-</td>
<td>+ maize, maize:cms-C, tobacco, <em>Petunia</em> (26-29)</td>
<td></td>
</tr>
<tr>
<td>NAD;Q₁ complex</td>
<td>Subunit 1</td>
<td>nad1</td>
<td>+</td>
<td>-</td>
<td>+ tobacco, maize (28) watermelon (30)</td>
</tr>
<tr>
<td>Subunit 3</td>
<td>nad3</td>
<td>+</td>
<td>-</td>
<td>+ maize, wheat (31)</td>
<td></td>
</tr>
<tr>
<td>Subunit 4</td>
<td>nad4</td>
<td>+</td>
<td>-</td>
<td>+ <em>Petunia</em>, wheat (NS)</td>
<td></td>
</tr>
<tr>
<td>Subunit 5</td>
<td>nad5</td>
<td>+</td>
<td>-</td>
<td>+ <em>Oenothera</em> (32)</td>
<td></td>
</tr>
<tr>
<td>Ribosomal proteins</td>
<td>rps4</td>
<td>-</td>
<td>-</td>
<td>+ <em>Oenothera</em> (33)</td>
<td></td>
</tr>
<tr>
<td>rps12</td>
<td>-</td>
<td>-</td>
<td>+ maize, wheat (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rps13</td>
<td>-</td>
<td>-</td>
<td>+ maize:cms-C, tobacco; (28) <em>Oenothera</em> wheat, (34-35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rps14</td>
<td>-</td>
<td>-</td>
<td>+ broadbean (16)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Mitochondrial translation product that has been identified. NS- not sequenced
1.6 Plasmid-like DNAs Associated with Plant Mitochondria

In addition to the main mitochondrial DNA, many plant species are characterised by the presence of low molecular weight DNA elements often called 'plasmids' or 'plasmid-like' DNAs (Pring et al. 1977). Eighteen different plasmids have been reported as either circular (in maize, sugarbeet, frenchbean, broadbean, sunflower and sorghum) or linear (in Brassica species and maize) molecules and range in size from 1.3kb to 11.3kb (Erickson et al. 1989). Generally, mitochondrial plasmids contain sequences which are not homologous to the main mitochondrial genome and which replicate independently of it. The origin of these mitochondrial elements remains unclear, but linear mitochondrial 'plasmids' contain polypeptides covalently attached to their 5' termini, a feature which they share with bacterial and animal viruses where these proteins participate in replication of the plasmid DNA (Kemble and Thompson 1982, Rekosh et al. 1977, Hirochika and Sakaguchi 1982).

Four plasmids found in rice mitochondria contain short, conserved regions of homology to each other and to the 1.9kb and 1.4kb plasmids found in maize mitochondria suggesting a common origin for the molecules (Shikanai et al. 1989). Mitochondrial plasmids have been found to characterise some male sterile cytoplasms of maize, rice and sugar beet, although their association with the male-sterile phenotypes remains to be clarified (Smith and Pring 1987, Saleh et al. 1989, Thomas 1986). In sunflower, a 1.45kb plasmid (P1) found in several cytoplasmic variants, has also been implicated in male sterility as discussed further in section 1.10.2. The cytoplasms of maize and sunflower contain double-stranded RNAs which are not homologous to either the main genome or to the DNA plasmids which are also present (Brown G. et al. 1986, Sisco et al. 1984). In maize containing an S-type cytoplasm, these RNAs replicate autonomously of the mitochondrial genome and so have been termed RNA 'plasmids' (Finnegan and Brown 1986). The origins and coding functions of these unusual molecules in plant mitochondria remain obscure.

1.7 Promiscuous DNA

A remarkable feature of the plant mitochondrial genome of several species is the presence of integrated copies of portions of the chloroplast genome (Stern and Palmer 1984, Stern and Lonsdale 1982). Petunia mtDNA contains thirteen different regions, accounting for approximately 5% of the total genome, which show similarity, on the basis of hybridisation analyses, to the chloroplast DNA (Hanson and Folkerts 1989). In maize, this so-called 'promiscuous DNA' (Ellis 1982), includes homologies to the 16s rRNA gene (Stern and Lonsdale 1982) part of the large subunit of ribulose bisphosphate carboxylase (rbcL), (Lonsdale et al. 1983) and tRNA genes (section 1.5.2). A function for the 16 rRNA and rbcL gene products within the mitochondria seems unlikely and transferred tRNA genes probably represent the
only expressed 'promiscuous DNA' sequences within the mitochondria (Lonsdale 1987). It is of note that a truncated chloroplast arginine tRNA sequence has been found to encode the last 14 amino acid residues of a novel 13kDa protein found in CMS-T maize mitochondria (Dewey et al. 1986 and section 1.9.4).

The fact that transfer of sequences from the chloroplast to the mitochondrion appears to have occurred frequently has intriguing implications regarding the mechanism by which the exchange is achieved. In addition to cpDNA sequences, some plant mtDNAs have also been found to contain sequences homologous to the nuclear genome (Schuster and Brennicke 1987b, Whisson and Scott 1985), and the nuclear genome has been found to contain homology to both the mtDNA and cpDNA (Cheung and Scott 1989, Kemble et al. 1983). In higher plants it seems that only the chloroplasts have remained devoid of 'promiscuous DNA' sequences, perhaps because their relatively small, conserved genomes cannot tolerate insertions. Although transposed cpDNA sequences make some contribution to mtDNA complexity, they cannot totally account for variations in the size of mitochondrial genomes between even closely related species (Stern and Palmer 1984).

The mechanism by which genetic information is transferred between chloroplasts, mitochondria and nuclei remains to be elucidated. In mtDNA from Oenothera, an ORF showing some sequence similarity with retroviral reverse transcriptase proteins has been identified (Schuster and Brennicke 1987a). Within the Oenothera mitochondrial genome, sequences adjacent to this ORF show homology with chloroplast (the psbE-psbF loci) and nuclear (18S rRNA) genes. It has been postulated that these sequences may have originated from RNA transcripts which entered the mitochondria, were reverse transcribed to complementary DNA sequences and subsequently integrated into the mitochondrial genome. Homology with reverse transcriptase within the putative nad1 locus in the mitochondrial genome of broadbean was mentioned earlier (section 1.5.3). However the broadbean ORF shows no homology to the Oenothera ORF, beyond the regions common to all reverse transcriptase type proteins which suggests there may have been two separate evolutionary acquisitions of this sequence in plant mitochondria. Reverse transcription predicts that only sequences which are expressed at the level of RNA will be available for transfer. As chloroplast homologous sequences present in plant mitochondrial genomes include some which are not transcribed within the chloroplast, this indicates that reverse transcription, if it does occur, it is not the only mechanism of information transfer and that DNA mediated transfer may also exist (Lonsdale, 1987).
1.8 Expression of Mitochondrial Genes

Gene expression encompasses all the processes involved in transcription of a DNA sequence into a complementary RNA and its subsequent translation into the protein which it encodes. RNA polymerase enzymes recognise specific 'promoter' sequences which determine the point where transcription is initiated on the DNA template. Ribosomal RNAs (rRNAs) are the structural components of the protein synthesis machinery and transfer RNAs (tRNAs) deliver specific amino acid residues to the ribosome for their polymerisation in the manner specified by messenger RNA (mRNA) molecules. The expression of genes as functional RNA or protein molecules can be regulated at several points in the pathway leading to their synthesis. The following section briefly describes aspects of mitochondrial gene expression in various organisms. Two systems in higher plants, thought to be a direct consequence of abnormal gene expression which leads to mitochondrial malfunction, are described in the final section.

1.8.1 Mammalian Mitochondrial Gene Expression

Transcription processes in the small, compact mammalian mitochondrial genomes are relatively simple and therefore well characterised (Attardi et al. 1985, Hixon and Clayton 1985). The majority of genes are encoded on the same DNA strand and are transcribed as polycistronic RNA transcripts that can be the length of the entire genome. Transcription is initiated from two divergent promoters to generate large precursor molecules which are cleaved to mature mRNAs, at tRNA sequences, located between protein coding sequences which act as processing sites. The site for transcription initiation, is located between the genes encoding the proline and phenylalanine tRNAs, as part of the control region which also contains the sequences involved in regulating mtDNA replication (Foran et al. 1988).

1.8.2 Yeast Mitochondrial Gene Expression

i. Transcript initiation.

In yeast (Saccharomyces cerevisiae) there are approximately twenty sites of mitochondrial transcription initiation dispersed throughout the genome (Grivell 1989). By comparison of these region a nonanucleotide consensus sequence has been derived (shown below) which is believed to represent the yeast mitochondrial promoter in which the final adenine residue denotes the site of initiation of transcription (arrow) (Christianson and Rabinowitz 1983):

\[
\begin{align*}
A & \\
5' & \text{or} & T & A & T & A & A & G & T & A & 3' \\
T & \\
\uparrow & 
\end{align*}
\]

Less conserved sequences outside this consensus may also contribute to promoter activity in yeast. 'Strong' and 'weak' promoters can be defined by their
ability to compete for limiting amounts of RNA polymerase and differences of up to 20-fold have been reported (Mueller and Getz 1986). Sequences downstream of the initiation site are, at least partly responsible for this variation, there being a preference for A and T at positions +2 and +3 respectively in stronger promoters.

Many genes in yeast mitochondria are not immediately adjacent to sites of transcription initiation and are expressed as parts of larger primary transcripts from which they are subsequently processed. Transcripts which are cotranscribed accumulate to variable steady state levels and transcription 'attenuation' is thought to play a role in regulating their production (Mueller and Getz 1986). Attenuation is caused partly by a slowing of the rate of transcription, with increasing distance from the point of initiation, so that more RNA copies of the first gene in the unit accumulate. Attenuation also occurs in mammalian mitochondria leading to the accumulation of ribosomal and some tRNA transcripts relative to those encoding protein genes (Clayton 1984).

ii. Transcript termination.

The 3' ends of many protein coding mRNAs in yeast have been found to coincide with a conserved dodecamer sequence shown below:

\[
5' \text{ AAUAUAUUUCCU 3'}
\]

In \textit{S.cerevisiae} this sequence is implicated primarily as a processing site, as a point of endonucleolytic cleavage within multigenic transcription units to generate mature mRNAs (Osinga \textit{et al.} 1984). It may confer stability to the 3' termini of such processed transcripts while further processing of the 5' ends occurs.

iii. Transcript processing.

Generally, expression of mitochondrial genes in yeast involves some form of transcript processing of the 5' and 3' noncoding sequences. After cleavage from a primary transcript, flanking regions are trimmed and intervening sequences, if present, are spliced out. In the case of tRNAs, post-transcriptional base modifications are necessary before a functional secondary structure can be assumed. Several yeast mitochondrial genes are interrupted by intervening sequences, the removal of which must be accurate and efficient and can represent a regulatory point in the production of mature mRNAs. Introns in several yeast mitochondrial coding sequences include ORFs which encode proteins required for splicing and maturation of their own or other gene transcripts (Grivell 1989).

iv. Nuclear control of mitochondrial gene expression.

Yeast cells can survive exclusively by fermentation and so can tolerate mutations which lead to defects in the respiratory pathways. Such mutations can arise in either the mitochondrial or the nuclear genes encoding structural components of the electron transport chain. It is also now evident that many nuclear gene products are involved in the regulation of expression of the yeast
mitochondrial genome through specific nuclear-mitochondrial interactions. Several yeast nuclear mutations have been isolated which specifically block the expression of only one or a few mitochondrial genes and indeed cases where several nuclear mutations produce the same effect on a single mitochondrial gene have been described (Michaeliset al. 1988). All of the nuclear 'factors' analysed so far act on post-transcriptional steps in the expression of their mitochondrial target genes and certain mitochondrial genes appear to have a set of nuclear functions involved in regulating their expression (Ben-Asher et al. 1989).

Fox et al. (1988) have isolated at least three nuclear genes PET494, PET54, and PET122, the products of which are required for translation of the coxIII gene. The site of action of these proteins lies within the 5' untranslated region of the coxIII mRNA, at least 200bp upstream of the translation initiation codon. They are thought to associate in some form of complex which functions as a coxIII specific translation initiation factor (Fox et al. 1988). The transcript level of PET 494 is reduced in yeast cells grown on glucose, conditions known to lower the level of the COXIII subunit and thus PET494 expression is controlled in a manner consistent with its role in modulating expression of the mitochondrial coxIII gene.

The product of the yeast nuclear gene cpb1 has been found to affect the level of COB protein which accumulates in yeast mitochondria (Rödel 1986). The protein encoded by cpb1 interacts with the leader sequence present upstream of the cob gene coding region to activate translation of the corresponding mRNA. The cpb1 gene has recently been mapped on chromosome IV of the yeast nuclear genome, adjacent to the gene for another mitochondrial protein, subunit 9 of the cytochrome c oxidase complex (Forsbach et al. 1989). No co-regulated transcription from these linked genes occurs but a specific increase in the level of transcription of cpb1 is reported under conditions of high O2. Such environmentally mediated regulation of expression of nuclear genes whose products control mitochondrial gene expression may be partly how the coordination between the genetic systems in the two compartments is achieved.

1.8.3 Plant Mitochondrial Gene Expression

Our understanding of transcription processes in plant mitochondria has been limited by the lack of a system for in vitro transcription to allow the analysis of promoter functions. However, observations based on sequence comparisons and transcript mapping studies have allowed descriptions of several characteristics of plant mitochondrial gene expression. Many plant mitochondrial protein coding genes are expressed as single, monocistronic mRNAs but evidence is accumulating that a number of genes exhibit more complex patterns of transcription. Some genes show multiple-sized transcripts some of which are much larger than necessary to encode the corresponding polypeptides. Such genes include cob, coxII and atpA in maize and atp6 and atp9 in tobacco (Dawson et al. 1984, Fox and Leaver 1981, Isaac et
al. 1985a, 1985b, Bland et al. 1986, 1987). The complexity of these transcriptional patterns is largely accounted for by heterogeneity in the 5' termini which appear to result from multiple transcript initiation events (reviewed by Levings and Brown 1989). Processing of 5' and 3' non-coding sequences from longer transcripts prior to translation also accounts for some of the complex transcript patterns observed. In the few plant mitochondrial genes which contain introns (coxII and the nad genes), transcripts corresponding to unspliced precursors and excised intron sequences contribute to the apparent complexity in several species (Fox and Leaver 1981, Bonen et al. 1984, Wissinger et al. 1988).

i. Transcript initiation.

Attempts to define those sequences which act as promoters for plant mitochondrial transcription have involved transcript mapping studies aimed at distinguishing primary (i.e. initiated from the DNA template) transcripts from those generated by processing. This is made possible by using the enzyme guanosyl transferase to specifically label unprocessed transcripts at their 5' terminal diphosphate or triphosphate group, in a process known as "capping" (Christianson and Rabinowitz 1983). The 5' termini of the RNA products of transcript processing are monophosphorylated and therefore cannot act as substrates for the addition of $^{32}$P labelled GTP catalysed by this enzyme. Termini of maize mitochondrial transcripts labelled in this manner map to two sites upstream of cob, three sites upstream of coxIII and as many as six sites upstream of atp9 (Mulligan et al. 1988a).

The 5' termini of 'cappable' transcripts corresponding to several plant mitochondrial genes have been found to map to sequences which show similarity to the consensus promoter sequence in yeast. For example, the longest transcript of one copy of atp9 in Petunia maps to the sequence 5' ATATAGTA 3' which is nearly identical to the yeast promoter (Rothenberg and Hanson 1987). This and other similar observations have led to the proposal of a longer plant mitochondrial promoter motif, having the consensus 5'-AAATYTCNTAAGTGAA-3' where Y denotes G or C and N denotes A, G, C, or T (Isaac et al. 1985b, Hiesel and Brennicke 1985, Young et al. 1986). However, transcript termini have been found to map to different points within this sequence, and it is not present upstream of all plant mitochondrial genes. In fact, there are a number of different sites to which mRNAs capped in vitro map, so that a general plant mitochondrial promoter cannot be defined. This has led to the suggestion that RNA polymerase recognition sites may lie at some distance from the initiation sites or that the enzyme may interact relatively non-specifically with its template to effect transcription initiation (Mulligan et al. 1988b).

ii. Transcript termination.

Little is known of the mechanism of transcript termination in higher plant mitochondria. Although the 3' termini of plant mitochondrial transcripts are less heterogeneous than the 5' termini, in a number of cases, multiple sites of transcript
termination may contribute to the observed complexity (Newton 1988). Plant mitochondrial mRNAs are not polyadenylated and sequence comparisons suggest that structural motifs at the 3' end of genes may be involved in the termination process. A comparison of the transcripts of the *Oenothera atpA* and *coxII* genes has shown that the last fifty nucleotides which they contain are identical and homologous sequences are also found in the region where several maize mitochondrial transcripts are thought to terminate (Schuster *et al.* 1986). Characteristically, these 3' regions contain inverted repeat sequences with the potential to form stem-loop secondary structures which may function as processing signals or termination sites in the transcripts where they occur. However, inverted repeats and predicted stem-loop configurations are not associated with all mitochondrial gene transcripts even within individual species (Hiesel *et al.* 1987, Bland *et al.* 1986, Rothenberg and Hanson 1987). Inverted repeats characterise the 3' termini of some chloroplast gene transcripts and appear to be involved in the processing of precursor mRNA and/or stabilisation of the mature message (Stern and Gruissen 1987).

### Transcript Processing

For some plant mitochondrial genes, RNA processing contributes to the complexity of the transcriptional patterns. Single transcript initiation sites for the *rrn26* and the *rrn18--rrn5* genes in maize occur approximately 200bp upstream of the 5' terminus of the 26S and 18S rRNA molecules. The primary rRNAs undergo 5' end processing and the 5S rRNA forms a dicistronic precursor with the 18S which undergoes further processing to release mature transcripts corresponding to each rRNA (Mulligan 1988b). A number of genes which show complicated transcript patterns in plant mitochondria have been found to contain regions showing conservation of 50-70 nucleotides at their 5' termini (Schuster and Brennicke 1989). These sequences, found upstream of both rRNA and protein coding transcripts from several different species, have been proposed as recognition sites involved in the processing of such transcripts. The distance between the processing consensus proposed by Schuster and Brennicke (1989) and the putative translational start site of the genes in which it is found, is variable but of the order of 100-300 nucleotides. The sequences are distinctly different from processing sites, proposed on the basis of sequence alignments to occur in the shortest transcripts of the *Petunia atp9*, maize *coxI* and wheat *coxII* genes (Young *et al.* 1986, see also Chapter 4).

The plant mitochondrial genes which show complex, interrupted coding sequences, such as *coxII*, *nad1* and 5 are initially transcribed as mRNA precursors (pre-mRNAs) from which the introns must be processed (Michel and Dujon 1983, Wissinger *et al.* 1988). Plant mitochondrial introns are group II type intervening sequences which are removed by a two-step pathway. This involves the formation of unusual secondary structure called a lariat produced by the binding of the
terminal phosphate of the intron at the 5' splice site with an adenosine residue lying within the intron near the 3' splice site (Brown J. et al. 1986). The initiation and regulation of pre-mRNA splicing in plant mitochondrial genes has not been studied in detail and the enzymes involved (whether nuclear or mitochondrially encoded) remain to be characterised.

iv. Co-transcription.

Although cotranscription of linked genes is generally not the case in higher plant mitochondria, it does occur. As described above, rnr18 and rnr5 genes are cotranscribed in all species examined to date. In wheat mtDNA, the initiator methionine tRNA lies immediately upstream of the rnr18 and is probably part of the same transcriptional unit (Gray and Spencer 1983). In Oenothera, the nad5 gene begins 380 nucleotides downstream of the rnr18-rnr5 region and is included in the same transcription unit. Transcriptional attenuation may account for the greater accumulation of the rRNAs relative to the nad5 mRNA (Wissinger et al. 1988).

The genes for nad3 and rps12 are cotranscribed in maize, wheat and Oenothera (Gualberto et al. 1988, Schuster and Brennicke 1987c) and in tobacco the nad1 and rps13 are cotranscribed with the atp9 gene which is situated immediately upstream (Bland et al. 1986). Recombination within the mitochondrial genome of CMS Petunia has resulted in the inclusion of the nad3 gene in a transcription unit containing a novel ORF which is implicated in generation of the sterile phenotype (section 1.9).

1.8.4 Transcripts with Unassigned Functions in Plant Mitochondria

The number of assigned reading frames within the plant mitochondrial genome still falls short of the number of polypeptides synthesised by the organelle in vitro. Attempts are being made to discover which regions, other than those for defined genes, are expressed in plant mitochondria. In a study of the 218kb genome of Brassica campestris, Markaroff and Palmer (1987) found that 61kb was transcriptionally active and could be represented on a genomic transcriptional map. Twenty-four abundant transcripts were detected of which ten could be accounted for by known mitochondrial genes (coxI, II and III, cob, atp6, atp9, atpA, rps13, rnr26, rnr18-rnr5). The additional transcripts remain to be identified. If they all represent translated RNAs, this would bring the number of polypeptides synthesized in the mitochondria to twenty-two, in good agreement with predictions from investigations of in vitro mitochondrial protein synthesis.

As stated earlier, very few of the proteins synthesised by plant mitochondria have been identified. The remaining polypeptides can be characterised by several different methods: (1) by placing the sequences for known genes and unassigned reading frames (URFs) into bacterial expression systems to elicit production of the corresponding proteins in vitro or (2) by synthesis of oligopeptides corresponding to short regions of the predicted amino acid sequences. Both of these can then be used to raise specific antibodies for selective immunoprecipitation with
mitochondrial translation products; (3) by purification of mitochondrial enzyme complexes followed by the production of antibodies raised against individual subunits and immunoprecipitation as above.

In addition to unassigned reading frames (URFs) which encode normal proteins, other classes of 'gene' have been found in the mitochondrial genomes of several plant species. Chimeric genes are sequences which have arisen by intra and/or inter-molecular recombination within the mitochondrial genome. In some cases, the resulting rearrangements have generated novel open readings while in others they have extended the open reading frame or modified the flanking sequences of existing genes (Dewey et al. 1986). Expression of these chimeric regions requires that they are located downstream of suitable promoter and translational control sequences as has been shown to occur in some CMS systems (discussed further in section 1.9).

Pseudogenes which lack promoter sequences or contain point mutations within their coding regions that prevent transcript initiation or cause premature termination of translation have also been found in plant mtDNA (Schardl et al. 1985, Schuster and Brennicke 1986). In addition, DNA sequences which are transcribed into RNAs large enough to encode polypeptides but which lack a continuous ORF have been identified (Hansen and Marker 1984). A reappraisal of the significance of such sequences, previously regarded as non-functional, may be necessary given the recent discovery that a process known as RNA editing occurs in plant mitochondria.

1.8.5 RNA Editing in Plant Mitochondria

RNA editing is defined as any process which results in a change to the nucleotide sequence of an RNA molecule from that of the DNA template which encodes it (Simpson and Shaw 1989). Such unfaithful transmission of genetic information was first described in the mitochondria of kinetoplastid protozoa such as Trypanosoma brucei (Benne et al. 1986). Long stretches of the nucleotide residue uridine are inserted into specific regions of the RNA transcripts from a number of so-called 'cryptic' mitochondrial genes while in others modifications, such as deamination of cytidine (C) to uridine (U) occur at selective base positions. The insertion, or in some cases deletion, alter the reading frame within the transcript to create translatable messages. C to U conversions have now been reported for the transcripts corresponding to several mitochondrial genes in wheat (Covello and Gray 1989, Gualberto et al. 1989).

It has previously been proposed that in plant mitochondria there is a departure from the universal code, with CGG specifying tryptophan instead of the usual arginine (Fox and Leaver 1981). This proposal was due to the frequent finding of CGG codons in plant mitochondrial sequences at positions where tryptophan is the conserved residue in proteins of other organisms. However, if sequences
containing C residues are edited to U residues in the mRNAs of some plant mitochondrial genes they will indeed encode tryptophan according to the universal code and thus explain the apparent coding anomaly.

Gualberto et al. (1989) have deduced mRNA sequences corresponding to selected regions of the wheat mitochondrial genome and have found that three UGG codons in the coxII gene (amino acid positions 57, 87 and 129) and two in the cob gene (amino acid positions 120 and 239) occur at positions occupied by CGG codons in the corresponding genomic sequences, and which represent conserved tryptophan residues in other organisms. Similarly, a CGG located at a position corresponding to a conserved tryptophan in the coxIII gene is specifically altered to a UGG in the mRNA transcript. In contrast, CGG codons in both the coxIII gene and another mitochondrial gene, rps12, at positions corresponding to conserved arginine residues remain CGG codons in the mRNA sequences.

By reverse transcription sequencing, Covello and Gray (1989) have also examined the sequence of the coxII mRNA in wheat and report a total of 8 C to U changes in this coding region when compared to the genomic sequence. Two of the proposed C to U conversions agree with those found by Gualberto et al. (1989) predicted to give conserved tryptophans at amino acid positions 87 and 129. The additional C to U conversions found by Covello and Gray predict changes of proline to leucine (position 188), serine to leucine (positions 196, 207 and 213) arginine to cysteine (228) and threonine to methionine (235). These conversions have the effect of altering the predicted wheat coxII amino acid sequence towards that of non-plant sequences at highly conserved positions.

Additional RNA editing processes are likely to be discovered and indeed, a specific insertion, rather than a conversion of a non-genomically encoded U, at the junction of two exons of the wheat nad4 transcript has been found (Gualberto et al. 1989). Complex editing, similar to the extensive processing of transcripts in several kinetoplastid genes, may yet be found in plant mitochondria. The discovery of the phenomenon in plants opens up yet another avenue of investigation into the intricacies of mitochondrial gene expression and the challenge of understanding the enzymatic mechanism and molecular selectivity of the editing process.

1.8.6 Translation of Plant Mitochondrial Genes

As the mRNAs encoded by higher plant mitochondria are seemingly much longer than the genes they contain, some mechanism must exist to confer specificity for translation initiation at the correct point in the mRNA. In eubacteria and chloroplasts RNAs it is thought that specificity for translation is conferred by a ribosome binding site which functions in binding with a complementary sequence (the Shine-Dalgarno sequence 5'CACCUCUCU3') located at the 3' end of the 16S rRNA (Shine and Dalgarno 1974, Whitfield and Bottomley 1983). A comparison of the 18S rRNAs in wheat (Spencer et al. 1984) and maize (Chao et al. 1984) shows
some similarity to the Shine-Dalgarno sequence (5'UGAAUCCU3') which could facilitate mRNA-ribosome binding in plant mitochondria. Analysis of several plant mitochondrial genes shows that most appear to be preceded, around 12-20 nucleotides, 5' to the AUG initiation codon, by a short sequence showing complementarity to the 18s rRNA 'Shine-Dalgarno type' sequence (AUUGAA) (Dawson et al. 1984). Some genes show additional complementarity with the 18S rRNA sequences, but a general ribosome binding site homology is not well conserved between plant mitochondrial genes (Boer et al. 1985).

Determining exactly which nucleotides function in translation initiation in different plant mitochondrial genes will await functional characterisation and mutational analysis of the regions immediately 5' to the AUG codons in different sequences. In plants, as in yeast, nuclearly encoded trans-activating factors (proteins which interact with DNA sequences and influence expression from them) may be required for translation of specific mitochondrial proteins or for post-translational modifications to allow their assembly into larger complexes. In relation to plant mitochondrial sequences there is, as yet, little understanding of how such trans-activating factors could be specifically involved in the regulation of organelle genome expression required for mitochondrial biogenesis and function.

To summarise the preceding sections, plant mitochondrial genomes are exceptional in their size although apparently not their coding capacity and we can conclude that some of the 'extra' DNA is concerned with the regulation of transcription from the coding regions. Generally, genes which have been localized on physical maps are encoded on both strands of mtDNA, scattered around the genomes in an order which varies even between closely related species and which probably reflects extensive genomic reorganisation. Explanations for the generation of variable mtDNA organisations invoke either de novo recombination or amplification of pre-existing recombinant forms of the genome. Recombination contributes to genome diversity by creating novel linkage groups and may have functional significance as a mechanism of dispersing regulatory sequences around the genome. Thus, the tendency of plant mitochondrial genomes to undergo rearrangement is a source of molecular heterogeneity which begs investigations of how the recombination system(s) are themselves regulated during plant development.

The inability of higher plants to survive without mitochondria has hampered a rapid genetic characterisation of the system. Most mutations in plant mtDNA will be either 'silent', i.e. unexpressed due to an alteration in a non-coding region of the genome, or lethal, due to mutation in an essential gene. Only two systems in higher plants have provided the opportunity to investigate the underlying cytoplasmic mutations. For the maternally inherited phenotypic traits of cytoplasmic male sterility (CMS) and non-chromosomal stripe (NCS) the evidence is convincing that the causal mutations reside in the mitochondria.
1.9 Phenotypic Mutations associated with Mitochondrial DNA Alterations

1.9.1 Non-Chromosomal Stripe Mutations in Maize

Non-chromosomal stripe (NCS) in maize is a maternally inherited mutation which results in reduced growth and abnormal leaf development. Mutants show variable leaf striping caused by alternate regions of normal green tissue and mutant pale tissue (Newton and Coe 1984). Such 'sectoring' also occurs in the kernels and is consistent with somatic sorting out of wildtype and mutant mitochondria and a resultant pleiotropic effect on the chloroplasts. Different isolates of the maize NCS phenotype contain specific alterations in the organisation of their mtDNA when compared to their parental cytoplasms. Although the nuclear genotype appears to influence the occurrence of NCS mutations, they are strictly maternally inherited thereafter (Coe 1983).

Two abnormal growth mutations, NCS2 and NCS3, are the consequence of rearrangements in, as yet, undefined mitochondrial genes. In the case of NCS2, reduced synthesis of a 24kDa mitochondrial polypeptide occurs and appears to cause a reduction in the level of chloroplast photosystem II core complex (Rousell et al. 1988). The region of mtDNA which is altered in NCS2 mutants identifies an aberrant set of mitochondrial RNA transcripts within affected plants when compared to normal maize, none of which correspond to known mitochondrial genes (Feller and Newton 1987). Another mutation, NCS5 has been isolated in which mtDNA isolated from the pale sectors of affected leaves, contain a deleted form of the coxII gene in contrast to the normal form of the gene present in the mtDNA from the green sectors (Newton et al. in press). The mutant form of coxII is not transcribed and presumably mitochondria which contain this gene lack normal cytochrome oxidase activity.

1.9.2 Cytoplasmic Male Sterility

Male sterility in higher plants can be simply classified as genetic (occurring either spontaneously or induced) or non-genetic (induced by chemical, physiological or ecological conditions). Inherited male sterility can be further classified into: (i) genic male sterility, which is controlled by nuclear genes whose action is not influenced by the cytoplasmic type and for which the pattern of inheritance is Mendelian, (ii) cytoplasmic male sterility which is controlled by specific sterility-inducing cytoplasms and inherited maternally and (iii) gene-cytoplasmic male sterility which results from the interaction of particular nuclear genes with certain types of cytoplasm (Kaul 1988). Cytoplasmic male sterility (CMS) has been described in over 150 plant species and is recognised, phenotypically by a failure of the mature plant to produce functional pollen but does not usually affect female fertility (Laser and Lerstern 1972). CMS occurs either spontaneously, in natural populations, or is generated in breeding programmes through interspecific and intraspecific crosses. The phenotype appears to arise from
an incompatibility produced when the nucleus of one race or species is combined with the cytoplasm of another.

The specific mechanism responsible for CMS differs among species and indicates that different factors are responsible for the trait in different species. In tobacco, for example, one CMS type is caused by stamens which become petaloid instead of developing into normal stamens (Kaul 1988). In contrast, pollen abortion at various stages of development occurs in CMS varieties of many species, including sunflower. In theory, any genetic element of the cytoplasm could be responsible for maternally inherited male sterility, indeed 'plasmid-like' DNAs have been implicated in several different species, including sunflower (Crouzillat et al. 1987, Turpen et al. 1988).

There is now a range of circumstantial evidence to support the contention that CMS results from mutations within the main mitochondrial genome in a number of plant species. Restriction endonuclease digestion patterns of mtDNA from fertile and sterile cytoplasms often show characteristic alterations, whereas chloroplast genomes from the same cytoplasms appear to be unaltered. Although the CMS determinants apparently reside in the mitochondrial genome, the expression of the phenotype is influenced by the nuclear genotype. This suggests that nuclear genes of a particular species may have co-evolved with those of the mitochondria and that specific interactions are involved in the efficient control of the mitochondrial gene expression at all stages of plant development.

1.9.3 The use of CMS in plant breeding

CMS is exploited commercially for the production of F₁ hybrids in maize, sorghum, rice, sugar beet, onion, sunflower and a number of other species (Kaul 1988). The phenotype prevents self-fertilisation of the maternal line, thus eliminating the need for labour intensive hand-emasculolation to promote cross fertilisation. The large scale use of CMS was made possible by the discovery of male fertile lines carrying specific dominant nuclear genes which suppress the CMS phenotype in the hybrid progeny. These genes (Rf genes) lead to the restoration of pollen fertility, which is essential when the commercial importance of a crop depends upon the seed produced. The different male-sterility conferring cytoplasmic types can be classified, within a species, on the basis of the nuclear restorer genes which are required to restore the fertile phenotype. However molecular and biochemical criteria can also be used to distinguish between CMS cytoplasms within species and have been used to investigate the nature of the CMS mutation(s) in a number of important crop plants. Currently, the best understanding we have of CMS at the molecular level comes from investigations with maize and Petunia.
1.9.4 CMS in Maize

There are three major groups of CMS cytoplasms in maize called CMS-T (Texas), CMS-C (Charrua) and CMS-S (USDA). These cytoplasms are distinguished by the genetics of fertility restoration (Laughnan and Gabay-Laughnan 1983), mtDNA restriction profiles (Pring and Levings 1978) and characteristic polypeptides synthesized by isolated mitochondria (Forde et al. 1978, Forde and Leaver 1980).

i. CMS-T. Maize plants carrying the CMS-T cytoplasm are susceptible to the race T of the fungal pathogen *Cochliobolus heterostrophus* (Southern corn leaf blight). The T-toxin produced by this fungus relaxes the normally stringently regulated permeability of the inner membrane in mitochondria of CMS-T plants, resulting in a leakage of cofactors such as NAD and an inhibition of the oxidation of malate and succinate substrates (Matthews et al. 1979). The CMS-T phenotype and susceptibility to the toxin are closely linked since revertants to fertility are also resistant to the toxin (Gengenbach et al. 1977, Brettell et al. 1980). CMS-T mitochondria synthesize a novel 13kDa polypeptide which is not detected in mitochondria from fertile, CMS-C or CMS-S maize (Forde et al. 1978). The 13kDa protein is constitutively expressed in all organs of CMS-T maize and, although apparently associated with the mitochondrial membranes, its exact location is uncertain (Leaver et al. 1988, Levings and Brown 1989).

Analysis of tissue-culture generated revertants of CMS-T plants has revealed that they lacked a particular region of the mitochondrial genome, present as a 6.7kb XhoI restriction endonuclease fragment in 'T-type' mtDNA (Dewey et al. 1986). This fragment contains a chimeric region of the mitochondrial genome consisting of several coding sequences which are unassociated in the mtDNA organisation found in normal, fertile maize. At least seven independent recombination events are thought to have contributed to the generation of this chimeric locus specific to mtDNA in cms-T maize. Part of the chimeric region has been sequenced and designated TURF2H3 (Levings and Dewey 1988). TURF2H3 includes homology to the 5' region of the *atp6* gene, internal and flanking region of the 26S rRNA gene (functional copies of both genes are also present elsewhere in the T mtDNA genome) and a chloroplast-derived arginine tRNA gene.

Two ORFs are present within the TURF2H3 sequence; one of 345bp encodes the 13kDa protein and has been designated *T-urf13*. It consists of 88 codons of the untranscribed 3' flanking region of *rrn26*, 9 codons of unknown homology, and 18 codons with homology to the coding region of *rrn26*. A second ORF, located 77 nucleotides 3' to *T-urf13*, and potentially encoding a 25kDa protein, has also been sequenced (Dewey et al. 1986). This ORF (designated ORF25) is transcribed, but as yet a corresponding translation product has not been identified. The pivotal event in the evolution of *T-urf13* has been the duplication of a 5kb region, normally found upstream of the *atp6* gene at the start of the chimeric coding region, thus
providing the necessary promoter for its transcription (Fauron et al. 1987, Wise et al. 1987a).

In maize CMS-T plants which have reverted to fertility the T-urf13 genes have either suffered deletion or been modified in some way through recombination (Rottman et al. 1987). In several cases recombination between two 55bp repeats which are at least 5kb apart in the mitochondrial genome appears to be the basis of the deletion event which alters the T-urf13 configuration. In one revertant, named T4, a tandem 5bp repeat in the T-urf13 produces an inframe stop codon which truncates the 13kDa protein to a predicted 8.3kDa (Wise et al. 1987b).

The relationship between T-toxin sensitivity and the 13kDa polypeptide encoded by T-urf13 has been established in heterologous systems of bacteria and yeast. By the use of artificial gene constructs containing the Turf13 sequence, expression of the 13kDa protein has been found to confer T-toxin sensitivity to the membranes of both E.coli and yeast when they are grown on a non-fermentable carbon source such as glycerol (Dewey et al. 1988, Glab 1988). Thus, there is now considerable evidence that the presence of the Turf13 gene, and associated synthesis of the 13kDa protein is causally linked to toxin-sensitivity and the CMS-T phenotype in maize. Exactly how the 13kDa protein is involved in conferring male sterility remains to be shown. It seems likely that it has some affect on membrane function, although no specific association of the 13kDa protein with a particular membrane complex has been demonstrated. The level of the 13kDa protein is reduced in the presence of the nuclear restorer gene Rf1 suggesting there is a direct link between the level of this protein and sterility. The abundance of several Turf13 transcripts is reduced, and a novel transcript is detected in plants which contain the Rf1 gene implying restoration involves both transcriptional and translational alterations (Kennell and Pring 1989). As there is no evidence that the nuclear gene Rf2 affects expression of T-urf13, yet both Rf1 and Rf2 are required for restoration of fertility, there may be more to the molecular basis of the T-type sterility than the 13kDa protein.

ii. CMS-C. In maize plants containing the CMS-C type cytoplasm, the mtDNA also includes chimeric ORFs apparently generated by rare recombination events. The ORF corresponding to atp9 in CMS-C plants, has a coding region identical with that found in the genome of fertile maize, but contains an unusual 5' region resulting in altered atp9 transcription in this cytoplasm (Levings and Dewey 1988). Also, a single copy of the coding region of coxII is present in CMS-C and is fused to the amino terminal and 5' flanking sequences of the atp6 gene. The remainder of the atp6 gene is itself part of another chimeric ORF containing the 5' flanking region of atp9 and 441 nucleotides showing strong homology to the chloroplast genome (Levings and Dewey 1988). The chimeric coxII and atp6 genes are the only functional copies of these sequences present within CMS-C mtDNA.
When mitochondrial translation products of this cytoplasm are analysed, a 17kDa polypeptide synthesised by CMS-C maize replaces a 15.5kDa protein found in mitochondria from N, S and T cytoplasms although the coding sequence for the 17kDa has not been defined. Forde et al. (1978) have shown that the level of the 17kDa protein does not change in the presence of the CMS-C restoring gene (Rf4) suggesting that if its expression is linked to sterility, the restoration process does not involve a reduced accumulation of the protein as appears to be the case with the 13kDa.

iii. CMS-S. CMS-S cytoplasms are characterised by the presence of two linear plasmid-like DNAs as mentioned earlier (section 1.6). The S1 (6397bp in size) and S2 (5453bp) plasmids contain identical 208bp terminal inverted repeats (TIRs) which share homology to a pair of 186bp repeats adjacent to the coxl gene on the main genome. The homology acts as sites for recombination of the plasmids into the genome resulting in the production of linear mitochondrial chromosomes with S1 or S2 attached at one end (Schardl et al. 1984). The recombinations place coxl in multiple genomic environments but expression of the gene is unaffected (Isaac et al. 1985b). In the N cytoplasm of maize, homology to S1 and S2 are found adjacent to recombinationally active 5.27kb repeat regions on the main genome (Houchins et al. 1986).

In addition to the TIRs, S1 and S2 contain 1254bp of homology adjacent to one of the repeats (Palliard et al. 1985). This region of homology contains an open reading frame which has been called ORF2 (1017bp). The S2 plasmid also contains an ORF1 (3513bp) and the S1 plasmid contains an ORF3 (2787bp) and an ORF4 (768bp) which are transcribed (Traynor and Levings 1986). Transcripts from the S-plasmid ORFs have been tentatively designated as an RNA polymerase (for the ORF1 of S2) and a DNA polymerase (for the ORF3 of S1) which are probably concerned with plasmid maintenance and are unrelated to CMS (Kuzmin et al. 1987, 1988) The mitochondria from CMS-C maize synthesize a number of high molecular weight polypeptides (ca. 125kDa) that are not detected in mitochondria from fertile maize (Forde and Leaver 1980). By cloning a portion of the ORF1 sequence from S2 into an expression vector, and using the fusion protein produced to elicit antibodies, it has been shown to selectively precipitate a polypeptide of ca.125kDa synthesized by S, but not N, mitochondria (Manson et al. 1986, Zabala et al. 1987). Similarly, the product of the ORF3 of S1 corresponds to a 103kDa protein synthesized in CMS-S maize mitochondria, but also in those of restored CMS-S plants and induced fertile revertants (Zabala et al. 1988). The nuclear restorer gene (Rf3), apparently has no effect on the synthesis of these proteins. Indeed, the relationship between the presence of the S plasmids and the CMS-S phenotype remains unclear and as mitochondria of CMS-S maize, restored to fertility, may or may not retain S1 and S2 depending on the nuclear genotype it seems unlikely that they are causally involved in generating the sterile phenotype (Escote et al. 1985).
1.9.5 Cytoplasmic Male Sterility in Petunia

In Petunia a chimeric mitochondrial gene has been identified in sterile plants and in somatic hybrid plants regenerated after fusion of protoplasts from CMS and normal fertile plants (Boreshore et al. 1985). Sequence analysis of this gene (designated Pcf) has shown it to be product of recombination between sequences from the atp9 and coxII genes and a third unidentified reading frame called urfS (Young and Hanson 1987). The Pcf locus, potentially encodes a continuous ORF of 38kDa, consisting of the first 35 codons of the atp9 gene, 159 codons of the coxII coding region plus the carboxy-terminus and 3' flanking region of the urfS sequence (see Figure 1.4). The 5' flanking region of Pcf contains the upstream region of atp9, including the putative site of transcript initiation associated with normal expression of this gene.

Complete, transcribed copies of the atp9 and coxII genes are also present in CMS Petunia. The normal coxII gene contains a large intron however in the Pcf gene, homology to the entire intron is missing (Hanson et al. 1988b, Figure 1.4). The urfS sequence which is fused to the coxII homology in Pcf, consists of 471bp showing no homology to known coding regions and includes four tandem duplications of a 54bp sequence. The 3' terminus of the Pcf transcript maps over 1kb downstream of the end of urfS. At a distance of 316bp downstream of the Pcf stop codon there is an 118 codon ORF encoding the nad3 gene of this genome. This nad3 sequence contains a complete coding region, but the 5' non-coding sequence diverges from that of the nad3 gene in fertile Petunia, a distance of 117bp upstream of the start codon due to the presence of urfS. This is the only nad3 gene in CMS Petunia and its expression is dependant upon cotranscription with urfS (Hanson et al. 1988a).

A transcript corresponding to the Pcf gene shows a 4-5 fold increase in the RNA from anthers at an early stage of development in CMS Petunia. In contrast, the atp9 transcript, which initiates from an upstream sequence identical to the Pcf promoter, shows no such increase (Young and Hanson 1987). No difference in the steady state level of Pcf transcripts was found between CMS plants and those restored to fertility, suggesting that any restoration function must act post-transcriptionally if the Pcf gene is responsible for the CMS phenotype. A novel 25kDa protein, synthesized within the mitochondria, has been detected in CMS Petunia and the level of this protein is increased in anther tissue relative to the level detected in the rest of the plant. Also, the level of the 25kDa polypeptide is found to be reduced in plants restored to fertility (Hanson et al. 1988).

Several hypotheses as to how the Pcf locus may be involved in male sterility in Petunia have been forwarded. If the Pcf transcript encodes the 25kDa polypeptide, it will contain ATP9 and COXII protein homology that could compete with the normal proteins for binding sites in the respiratory complexes and lead to mitochondrial malfunction. An increased level of the variant polypeptide in the
anthers would result in greater competition with a concomitant reduction in the ability of the mitochondria to supply the demand for high rates of respiration associated with pollen development. Another possibility is that expression of the nad3 mRNA is incorrectly regulated due to the Pcf fusion. This may introduce an imbalance in the level of gene products for complex I which becomes critical in the anthers causing respiratory deficiency leading to abortion of normal pollen differentiation.

**Figure 1.4 Organisation of the Pcf gene in CMS Petunia**

The chimeric Pcf gene found specifically in CMS Petunia (S) and the sequences contributing to its organisation are shown. The 5' coding and flanking region of Pcf consists of an incomplete atp9 sequence (dark stippling). This is fused to a region of coxII exon homology (black) and an unidentified reading frame, urfS (spotted). A complete coding sequence for nad3 (light stippling) is present downstream of the urfS open reading frame. This gene diverges from the nad3 sequence present in fertile Petunia (F) in the 5' flanking region. Transcription of the nad3 gene in CMS Petunia occurs by co-transcription with the Pcf sequence.

1.10 Cytoplasmic Male Sterility in Sunflower

A stable source of genetic CMS in sunflower was first described by Leclercq (1969) as the product of an interspecific cross between *Helianthus annuus* L. (domesticated sunflower) and *H. petiolaris* Putt. Since then, over twenty other sources of sunflower CMS have been produced by interspecific and intraspecific crosses, however this original source is still used almost exclusively in breeding programmes around the world (see Figure 1.5). Such a lack of cytoplasmic diversity is an undesirable situation as was found for the USA maize crop in 1970, when over 80% of the F1-hybrid seed germinated contained the CMS-T cytoplasm which showed a greatly increased sensitivity to the fungus *Cochliobolus heterostrophus*
(Southern corn leaf blight). This led to a loss of ca. 10-20% of the corn crop at a cost in excess of 1 billion US dollars. This dependance on a single, uniform cytoplasm as a source of CMS highlighted the need to identify or create diverse CMS conferring cytoplasm for hybrid breeding programs.

1.10.1 Ultrastructural Observations of Microsporogenesis in Sunflower

The differentiation of the pollen grain and embryo sac is all that remain of the gametophytic stage in the alternation of generations of flowering plants. As the haploid male gametophytic generation, the pollen grain depends on the sporophytic diploid plant for its production and exists only until union with the female gametophytic embryo sac at fertilisation. Pollen differentiation occurs within the anther via a series of recognisable stages. Cells within the central sporogenous tissue of the anthers, called pollen mother initials or microsporocytes, undergo two meiotic divisions to generate haploid microspores. Each pollen mother cell produces a tetrad of microspores that undergo a number of surface modifications before being released as mature pollen grains (Dickinson 1987).

In sunflower, as in all angiosperms, a cell layer called the tapetum invests the developing microsporocytes. The role of the tapetum is presumed to be one of nutrition, acting as a route for the passage of assimilates to the microsporocytes. The tapetum appears to be a highly active tissue, containing rough endoplasmic reticulum associated with the synthesis of secreted proteins and mitochondria for providing energy and metabolic intermediates (Chapman 1987). In many plant species, including sunflower, aberrant changes in the tapetal cell layer are the first detectable deviation from normal development seen in CMS lines (Bino 1985a, 1985b, Lee et al. 1979).

Homer (1977) used the light and electron microscope to compare the development of anther and sporogenous tissue in a normal fertile line and a CMS sterile line. In sunflower, microsporogenesis can be divided into 11 stages, with stages 1-4 encompassing the differentiation of sporogenous tissue to the initiation of tetradst and stages 5-11 spanning the period from late tetrad development to the mature pollen. As far as stage 4, both fertile and sterile lines develop similarly. The tapetal cell layer of sunflower normally expands and invades the sporogenous tissue during the leptotene stage of the first meiotic division (stage 5). It then degenerates around the late tetrad stage and dissolution of its contents occurs to provide proteins and lipids which become incorporated into substances such as pollenkitt and tryphine which coat the pollen grains (Mascarenhas 1988). In CMS sunflower however, at stage 5 the tapetal differentiation becomes completely disorganised. Cells of the middle layer of the anther wall become highly enlarged and the tapetum expands earlier than normal, advancing so rapidly that it crushes the developing microspores which differentiate no further. Both the tetrads and the tapetum collapse and complete pollen abortion results (Horner 1977).
Figure 1.5 Production of hybrid sunflower using the CMS cytoplasm

A schematic representation of the creation and maintenance of cytoplasmic male sterile lines and their use in the production of hybrid sunflower is shown.

I Maternal (seed) and paternal (pollen) fertile lines are crossed to produce a hybrid line.

II The hybrid is backcrossed with the original pollen parent until male sterile progeny are obtained. These are then maintained by backcrossing (which may be with the original pollen parent).

III The male sterile can be restored to fertility by nuclear genes (Rf) within a restorer line, to produce a restored F1 hybrid.

Thus, CMS lines are hybrids in which a nucleus is maintained in a foreign cytoplasm in the absence of nuclear restorer genes.
1.10.2 Molecular Analyses of CMS in Sunflower

Preliminary molecular analysis of the molecular basis of CMS in sunflower concentrated on the presence of the 1.45kb ($P_1$) circular plasmid-like DNA found in fertile sunflower which was apparently lacking in sterile lines (Brown G. et al. 1986, Leroy et al. 1985). However, a subsequent survey of sixteen sources of sunflower cytoplasm has indicated the presence of $P_1$ or a related plasmid $P_2$ (1.8kb) in some sterile types thus negating the reported correlation with the CMS phenotype (Crouzillat et al. 1989). These plasmids are not homologous with the main mitochondrial genome, or with chloroplast DNA, but do hybridise with nuclear DNA suggesting they may be of nuclear origin. Sequence analysis of $P_1$ has revealed no large open reading frame although two low molecular RNA species, homologous to $P_1$, are detected in a fertile line (HA89B also known as 841B) (Perez et al. 1988). Various authors have concluded that neither $P_1$ nor $P_2$ is in any way associated with the CMS phenotype (Perez et al. 1988, Crouzillat et al. 1989).

During the course of this investigation, physical maps of the mitochondrial genomes of an isogenic couple of sunflower were published (Siculella and Palmer, 1988). The estimated size of the genomes in the fertile (HA89B) and sterile (HA89A) lines investigated were 300kb and 305kb respectively. A tripartite organisation (section 1.4.2) resulting from recombination across a 12kb direct repeat has been proposed, but the authors state that the maps do not account for small repeats (i.e. less than 500bp) which may be present in the genomes. Several of the known mitochondrial genes regions have been located on the maps and indicate that variation between genomes of fertile and sterile lines is restricted to a region containing the $atpA$ and $cob$ genes. A different isogenic couple from sunflower (fertile line 842B and sterile line 842A) was investigated in this project and will be discussed in comparison to these reported results.

1.10.3 Aims of this thesis project

Although largely circumstantial, the evidence implicating mutations in plant mtDNA as the basis for the CMS phenotype is now overwhelming. In those plant species examined to date, characteristic differences exist between the mitochondrial genome organisation of male-fertile and male-sterile lines. In several cases, specific recombination events result in unusual gene organisations leading to the creation of novel open reading frames or modification of existing genes, which, when expressed, lead to the synthesis of variant polypeptides. The presence of such aberrant proteins may be detrimental to normal mitochondrial function by influencing the structure, and/or activity of the major respiratory enzyme complexes. A reduction in mitochondrial efficiency may not be tolerated at all stages of the plant life cycle, and could become critical at stages dependant upon increased mitochondrial biogenesis and activity such as those associated with pollen development.
There is little doubt that, as in other species, the CMS phenotype in sunflower is due to mutation in the mitochondrial genome. Although studies have been initiated, there is, as yet, no understanding at the molecular and physiological levels of the mitochondrial genetic lesion which causes the sterile phenotype in this species. The aim of the project described in this thesis was to investigate the molecular basis of CMS in sunflower by a comparative analysis of the organisation and expression of the mitochondrial genome from fertile and sterile cytoplasms. Several approaches to achieving this aim were used in an attempt to gain a complete description of the molecular variation which exists between the two phenotypes. These included:

1) an investigation of the coding capacity and genome organisation present in mitochondria from fertile and sterile sunflower (Chapter three);
2) attempts to define specific variations in the organisation of the two genomes, which correlate with the sterile phenotype and a description as to how these may have arisen (Chapter four);
3) an investigation of transcription of the mitochondrial genome in fertile and sterile sunflower (Chapter five);
4) an examination of mitochondrial genome expression by analysis of the products of *in organello* translation to establish whether there are differences between proteins synthesised in mitochondria from fertile and sterile sunflower (Chapter six).

The ultimate aim of these investigations was to establish whether any variation found in mitochondrial genome organisation and expression could be causally related to the male-sterile phenotype. Chapter seven summarises the conclusions from the study and discusses how they can be integrated into a description of how CMS may have arisen in sunflower. Finally some proposals for future work are presented in the hope they will lead to a complete understanding of the molecular basis of the male sterile phenotype in this species.
CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Seed
All plant material was grown from seed kindly provided by Rhône-Poulenc Agrochemie of France. The varietal lines used, according to the company nomenclature, were as follows (phenotype given in brackets): RPA842A-Helianthus annuus (sterile); RPA842B-H.annuus (fertile); RPA843R-restorer line (fertile); RPA844H-restored hybrid (fertile). In the remaining text, these are described phenotypically or abbreviated to S, F, R and H respectively. The wild-type progenitor of 842A, Helianthus petiolaris was also used for the transcript studies as discussed in Chapter 5. Another sterile line, RPA841A and a fertile, RPA841B were used in the mitochondrial protein synthesis experiments described in Chapter 6. Seeds were grown in medium grade 'vermiculite' for etiolated plant material or Levington Fisons M3 potting compost for sunflowers grown to maturity in a light growth chamber (Fisons model 600G3/THTL).

2.1.2 Chemicals
Chemicals were supplied by British Drug Houses [BDH], Bethesda Research Laboratories [BRL], the Sigma Chemical Company [SCC], Difco Laboratories [DF], or Boehringer Mannheim [BM] unless indicated otherwise.

2.1.3 Enzymes
Calf intestinal alkaline phosphatase: BM
DNA polymerase I: Amersham International plc.
DNase (RNAase free): P&S Biochemicals Ltd.
Klenow (large fragment of polymerase I): BRL
Lysozyme: SCC
Proteinase K: BM
Restriction endonucleases: BRL (except Sau3A Amersham, AluI BM)
RNAase A (bovine pancreatic): BM
T4 DNA ligase: BRL
T4 polynucleotide kinase: Amersham International plc.
Sequenase (version 2): United States Biochemical Company
2.1.4 Radioisotopes

Radioactive isotopes, supplied by Amersham International plc., as stabilised aqueous solutions were: deoxycytosine 5'\([\alpha-32P]\) triphosphate (\(\alpha\)-dCTP), specific activity 110TBq (3000Ci)/mmol; deoxyadenosine 5'\([\alpha-35S]\) triphosphate (\(\alpha\)-dATP), >37TBq (>1000Ci)/mmol; deoxyadenosine 5'\([\gamma-32P]\) triphosphate (\(\gamma\)-dATP), 110TBq (3000Ci)/mmol. L-\([35S]\)-methionine (\(35S\)-met), specific activity 43.2TBq (1169Ci)/mmol was supplied by New England Nuclear Research Products (DuPont).

2.1.5 Film and Intensifying Screens

Cronex 4 X-ray film and intensifying screens (DuPont) were used for autoradiography. Ilford HP5 film and a Polaroid camera fitted with an orange-G filter (Hoya) were used for photographing agarose gels with a short wave ultraviolet transilluminator supplied by Ultraviolet Products Ltd.

2.1.6 Centrifuge Equipment

Sorvall RC-5B centrifuges were used in conjunction with Sorvall GSA fixed angle rotors (6 x 250ml polycarbonate bottle capacity) and Sorvall SS-34 fixed angle rotors (6 x 50ml, 30ml or 15ml Corex tubes) for isolation of mitochondria. The RC-5B plus a Sorvall HB4 swing-out rotor were used for large scale nucleic acid purification. A Sorvall OTD65B ultracentrifuge and an AH629 swing-out rotor (with 36ml polypropylene tubes) were used for sucrose gradients and with a Sorvall 70Ti rotor for mtDNA-cesium chloride (CsCl) gradients. A Beckman TL100 bench-top ultracentrifuge, a T100.1 fixed angle rotor and Beckman 3ml heat-seal tubes were used for purification of plasmid and cosmid DNA on CsCl gradients. Eppendorf or MSE Micro-Centaur fixed angle microfuges (12000xg max) were used to centrifuge 1.5ml polypropylene tubes (Treff) during all small scale manipulations.

2.1.7 Mitochondrial DNA clones

\textbf{i. Maize mtDNA clones.} The maize mtDNA gene probes used in this project were previously isolated by Dr. P.G. Isaac (M13 clones of the \textit{atpA} gene from N-type maize, M13 clone of the \textit{coxI} and an M13 clone specific for the intron of \textit{coxII}), Dr A.J. Dawson (M13 clones of \textit{cob}) and Dr V.P. Jones (M13 \textit{coxII} exon clone). A 2.7kb \textit{HindIII} fragment and a 2.2kb \textit{XbaI} fragment containing the \textit{atp6} and \textit{atp9} genes respectively were available as plasmid clones. A chemically synthesised oligonucleotide specific to the maize \textit{nad3} gene was obtained based on the published sequence (Gualberto \textit{et al.} 1988).

\textbf{ii. Oenothera mtDNA clones.} A 5.2kb \textit{BamHI} fragment containing the genes \textit{rrn18}, \textit{rrn5} and \textit{nad5}, an 800bp clone of \textit{coxIII}, and a 2kb \textit{HindIII} fragment containing homology to \textit{nad1} all from Oenothera were kindly supplied by Dr. Axel Brennicke, Berlin.
iii. Wheat mtDNA clone. An 280bp BglII-XhoI M13 clone internal to the \textit{rps}12 gene from wheat was a gift from Dr. J.M. Gualberto, Strasbourg.

2.1.8 Bacterial host strains and vectors
The \textit{E.coli} strains used as hosts for cloning were
\begin{itemize}
\item HB101 \hspace{1cm} (F',\textit{hsdS}20 [r',m'],\textit{recA}13, \textit{ara}-14, \textit{proA}2, \textit{lacY}1, \textit{galK}2, \textit{rpsL}20, \textit{xy}l-5, \textit{mt}-1, \textit{supE}44) (Maniatis \textit{et al.} 1982)
\item JM101 \hspace{1cm} (\Delta (\textit{lac pro}), \textit{thi}, \textit{Sup E}, \textit{F'}\textit{traA}36, \textit{proAB lac}iqz) \hspace{1cm} (Yanisch-Perron \textit{et al.} 1985)
\item ED8767 \hspace{1cm} (\textit{hsdR'}, \textit{recA})
\end{itemize}
The vectors which were used have all been described elsewhere, with respect to restriction maps and cloning procedures:
\begin{itemize}
\item bacterial- \textit{pUC}18 (Vieira and Messing 1982)
  \hspace{1cm} - \textit{pAT}153 (derivative of \textit{pBR}322, Twigg and Sheratt 1980);
\item bacteriophage - \textit{M13mp}18 and \textit{mp}19 (Yanisch-Perron \textit{et al.} 1985);
\item cosmid - \textit{pJB}8 (Ish-Horowicz and Burke, 1981)
\end{itemize}

2.1.9 Other materials
Electrophoresis tanks were from Pharmacia (horizontal electrophoresis models \textit{GNA} 200 and 100) or BRL (vertical sequencing tank) run on Pharmacia \textit{EPS} 500/400 power packs. The vacuum blot transfer apparatus for DNA gels was an LKB Vacugene and Hybond-N nylon membranes were from Amersham International plc. The vacuum gel drying apparatus was from Zabona, Basel. 3MM chromatography paper and GF/C filters were from Whatman Ltd. and DEAE membranes from Scheicher & Schull. "Miracloth" was from Novabiochem Ltd., muslin was from Turnbull and Wilson, Edinburgh and Saran wrap was supplied by the Dow Chemical Company.

Herring sperm DNA was sheared using a Soniprep 150 sonicator. DNA and RNA concentrations were determined by scanning in a UV-visible range photospectrometer (Varian) and autoradiographs or photographs of gels were subjected to densitometer tracing in a QuickScan densitometer (Helena Laboratories). Sunflower tissue was homogenised in a Polytron homogeniser from Kinematica. Light intensity in the growth chamber was measured with a \textit{LI}-185 Quantum radiometer from Lambda Instruments. Pippetmen were by Gilson, and tips by Treff (yellow) and BM (blue). Glassware was from Volac or Corex and plastic petridishes from Sterilin.
2.2 METHODS

2.2.1 Growth of Plant Material

i. Growth of Etiolated Sunflower Seedlings. Seed was weighed and surface sterilised by soaking in dilute sodium hypochlorite (1-1.4% available chlorine) [A &J Beveridge Ltd.] for 10 minutes (min.) at room temperature, rinsed several times with dH2O and then imbibed for 18-24 hours in running water (ca.16°C). 50 grams of seed was planted per 60 cm x 30 cm x 7 cm tray filled to a depth of 3 cm with medium grade 'Vermiculite'. The trays were well watered with dH2O, covered with silver foil and allowed to drain before being placed in a dark room maintained at 26°C-28°C. Etiolated cotyledons and hypocotyls were harvested after 5 days, rinsed in dH2O to removed residual growth medium, weighed and used in mitochondrial preparations.

ii. Growth Under Diurnal Light and Temperature Conditions. Seeds were planted 2-3 cm deep, four per 18 cm diameter pot, and placed in a growth chamber under conditions of 12 hour light (at 3.8 mw cm⁻²) at 26°C, 12 hour dark at 22°C and watered daily. Tissue from light-grown material was harvested directly into liquid nitrogen and kept on ice in preweighed beakers until use.

2.2.2 Purification of Sunflower Mitochondria

Harvested tissue was used for purification of mitochondria by a method based on that for maize described by Leaver et al. (1983) and the entire procedure was carried out at 4°C. Approximately 50 g of etiolated hypocotyl and cotyledon tissue were homogenised using several 5 second bursts with a Polytron homogeniser in 100 ml of ice-cold grinding buffer (0.4 M D-mannitol, 25 mM 3-[N-morpholino]propanesulfonic acid (MOPS), 1 mM ethyleneglycol-bis-N,N,N',N'-tetraacetic acid (EGTA) pH 7.2, [all from SCC], 0.1% w/v BSA [BM], 8 mM cysteine [SCC] pH to 7.8 with 5N KOH). The homogenate was filtered through four layers of muslin cloth, to separate any unground material and then passed through four "Miracloth" filters, to reduce contamination by nuclei. The homogenate was then centrifuged at 3000 x g for 3 minutes at 4°C to remove cell wall debris, nuclei and most of the plastids. The supernatant was rapidly decanted into fresh tubes and centrifuged at 10000 x g for 15 min. The resulting pellet was then gently resuspended in 25 ml wash buffer (0.4 M mannitol, 5 mM MOPS, 1 mM EGTA pH to 7.2 with 5N KOH) using a Teflon homogeniser in a volumetric glass tube. The supernatant from a second low speed centrifugation at 3000 x g for 3 min. was carefully decanted into fresh tubes and recentrifuged at 10000 x g for 15 min. producing a crude mitochondrial pellet. The pellet was resuspended in 2 ml wash buffer using a glass rod and slowly layered onto a 36 ml sucrose step gradient consisting of 0.6 M, 0.9 M, 1.2 M, 1.45 M and 2.0 M sucrose [BDH] in 10 mM N-tris(hydroxymethyl)methylglycine (tricine) [SCC] and 1 mM EGTA pH 7.2 which had been left overnight at 4°C to linearise.
Gradients were centrifuged at 4°C in a Sorvall AH629 swing-out rotor. This was accelerated slowly to 5000rpm then normally, to reach 20000rpm (equivalent to a maximal 90000xg) and centrifuged for 1 hour. At the end of the run the rotor was allowed to coast to a halt to ensure proper reorientation of the gradients. The buff-coloured band of mitochondria positioned at the 1.2M region of the gradient was recovered using a bent pasteur pipette and collected in a volumetric conical tube. The volume of the mitochondrial fraction was measured and transferred to a 30ml Corex tube before the addition dropwise, and with constant gentle shaking, of 1.5 times the volume of diluting medium (0.2M mannitol, 10mM tricine pH 7.2, 1mM EGTA pH 7.2). This slow procedure was necessary to prevent osmotic shock of the mitochondria which would occur if they were diluted too quickly. The diluted fraction was centrifuged at 10000xg for 15 min. and the pellet resuspended in resuspending medium (0.4M mannitol, 10mM tricine pH 7.2, 1mM EGTA pH7.2) for immediate use in amino acid incorporation experiments (see 2.2.17) or the appropriate lysis buffer as described below. Mitochondrial pellets to be used for nucleic acid purification could be stored, after quick freezing in liquid N₂, at -80°C at this stage.

2.2.3 Isolation of Mitochondrial DNA
Mitochondria purified through sucrose gradients were used to prepare mitochondrial DNA (mtDNA), by CsCl density equilibrium centrifugation as described by Fox and Leaver (1981). After dilution, mitochondria were resuspended in DNA lysis buffer -100mM Tris-hydroxymethylaminomethane (Tris)-HCl [BM] pH8.0, 100mM ethylene- diaminetetraacetic acid, disodium salt (EDTA) [BDH] containing 0.5mg/ml Proteinase K. N-lauryl-sarcosine (sarcosyl) [BDH] was added to a final concentration of 0.5%w/v and the total volume brought to 5.65m1s with lysis buffer. Tubes were capped with parafilm to prevent evaporation during a 60 min. incubation at 60°C in a shaking water bath. After incubation, 6.4g of CsCl [SCC] and 1ml of 700μg/ml ethidium bromide (EtBr) [SCC] were added and mixed with the sample. The entire solution was added to a 13ml capacity heat-sealable plastic tube, overlain with paraffin oil to fill and balance tubes, and centrifuged at 46000rpm (200000xg) for 20 hours at 12°C in a 70Ti rotor. MtDNA which had banded in the gradient was visualised under UV light and removed by puncturing the side of the tube with a fine-gauge needle and slowly drawing off the band into a 1ml syringe. EtBr was removed by repeated extraction with butan-1-ol [BDH] (pre-saturated with NaCl [BDH]).

Two volumes of TE8.0 (10mM Tris-HCl pH8.0, 1mM EDTA pH8.0) and 6 volumes of cold ethanol (95%) [James Burrough Ltd] were then added and the DNA precipitated overnight at -20°C. The precipitation was centrifuged at
10000xg for 30 min. in the HB4 rotor and the pellet resuspended in a total of 400μl of TE8.0 and transferred to a microfuge tube. A 1/10 volume of 3M sodium acetate (NaAc) [BDH] pH5.6 and 2.5 volumes of cold ethanol were added and the DNA reprecipitated at -20°C for 1 hour. After centrifugation for 10 min. in a microfuge, the pellet was washed with 70% ethanol and centrifuged for a further 10 min. before the DNA pellet was dried in a vacuum desiccator. The dry pellet was resuspended in a small volume of TE8.0 and left, on ice for 1 hour to rehydrate. MtDNA was stored at 4°C until further use.

A spectrophotometric scan from 320nm-220nm of an aliquot of the sample was used to estimate concentration (assuming that an absorbance at 260nm, i.e. an A\textsubscript{260} of 1.0 corresponds to an DNA concentration of 40μg) and purity of the preparation (using the ratio A\textsubscript{260}:A\textsubscript{280} to estimate the protein contamination).

2.2.4 Electrophoresis of Sunflower Mitochondrial DNA

The concentration, intactness and purity of mtDNA were checked by agarose gel electrophoresis. 1-3μl of mtDNA in TE8.0 was electrophoresed on a 'mini' 0.7% (w/v) agarose [BRL] gel containing 1 x TBE (89mM Tris-borate, 89mM boric acid [BDH]) running buffer. EtBr in the gel (final concentration of 0.5μg/ml) allowed the mtDNA to be visualised by fluorescence under UV-illumination and compared to phage lambda (λ) DNA standard size markers of known concentration. MtDNA was digested with restriction endonucleases enzymes, used in accordance with manufacturers recommendations for salt and temperature conditions. Typically, 3μg mtDNA was digested in a volume of 40μl containing 10 units of enzyme (where a unit is defined as the amount of enzyme required to completely digest 1μg λ DNA in 1 hour in the appropriate buffer). Varying concentrations of agarose, from 0.7-1.0%w/v, were used to give efficient separation of the restriction endonuclease generated fragments.

2.2.5 Mitochondrial DNA Blot Hybridisation

Following agarose gel electrophoresis mtDNA was transferred to Hybond-N nylon membrane by a method based on the traditional method of Southern (1975). DNA blots were produced using a vacuum blotting unit which has the addition of a vacuum pump to aid capillary transfer of nucleic acid from the gel to the filter surface. A piece of Hybond-N nylon filter was cut to the size of the gel and placed beneath the gel on a supporting tray in the unit. A few mls of 0.25M hydrochloric acid (HCl) [BDH] were pipetted onto the gel and allowed to soak in to introduce single-strand 'nicks' into the DNA and make subsequent transfer of high molecular weight fragments more efficient. DNA was then allowed to transfer from the gel to the filter for 10 minutes in a denaturing solution of 0.5N sodium hydroxide (NaOH) [BDH] and 1.5M sodium chloride (NaCl) and with a constant vacuum pressure of 30cm\textsuperscript{2}.H\textsubscript{2}O. Residual denaturing solution which had not soaked into the gel was removed before a
neutralising solution (3M NaCl and 0.5M Tris HCl, pH 7.0) was added for 10 minutes. Finally the gel was immersed in 20 X SSC (3MNaCl, 0.3M sodium citrate [Aldrich Chemical Company] pH7.2) and left for 1 hour. The gel was restained after the blotting procedure to ensure all the DNA had been transferred. The filter was removed and completely air-dried before being wrapped in ‘Saran-wrap' and placed on a UV-transilluminator for 3' to fix the DNA to the filter. Filters were stored in Saran-wrap at 4° C until further use.

For probe hybridisation, and unless otherwise stated, DNA filters were pre-hybridised, to minimise non-specific binding of DNA, in plastic bags containing 5xSSC, 1x Denhardts (0.02%w/v with respect to type 400 Ficoll [SCC], polyvinylpyrrolidone (PVP) [SCC] and BSA, 0.1%v/v sodium dodecyl sulphate (SDS) [BDH] and 250μg/ml final concentration of sonicated herring sperm DNA [Serva Feinbiochemica] which had been boiled and quickly cooled on ice to ensure single-strandedness. Prehybridisations were performed by placing sealed plastic bags, containing the filter and prehybridisation solution on a rocking table in a 65° C oven. Probes were added to the prehybridisations bags 2-20 hours after the start of prehybridisation and left at 65° C overnight. Following hybridisation, the probe solution was removed and the filter placed in a series of washes to remove non-specifically bound radioactivity.

The washes, increasing stepwise in stringency due to a reduction in the salt concentration, were first a 15 min. wash of 5xSSC, 0.1%SDSv/v at 65° C, a second 30 min. wash of 2xSSC, 0.1%SDSv/v at 65° C and a third 10 min. wash of 1xSSC at 65° C. Filters were checked with a radioactive monitor for the 'signal to background' ratio and further washes of 0.5xSSC and 0.2xSSC at 42° C for 30 min. performed as thought necessary. In some cases a high stringency wash of 0.1xSSC at room temperature was used. After washing, filters were air-dried, wrapped in Saran-wrap and autoradiographed by exposure to X-ray film in light-tight cassettes placed at -70° C. If a filter was to be reprobed, the first probe was stripped by washing in 0.4N NaOH at 45° C for 30 min. followed by a wash in 0.1xSSC, 0.1%v/v SDS, 0.2M Tris-HCl pH 7.5 under the same conditions. The filter was air-dried and then ready for reuse.

2.2.6 Preparation of Sunflower Mitochondrial RNA

Purified mitochondria were used to isolate mitochondrial RNA (mtRNA) for blot hybridisations (‘Northern' hybridisations). Precautions to minimise degradation by RNAase enzymes were taken: all glassware was sterilised by baking at 120° C for 8 hours; solutions were freshly made and autoclaved before use; clean surgical gloves were used and the entire preparation was done at 4° C. Firstly, mitochondria purified through sucrose gradients were resuspended in 1ml of NDS lysis buffer (0.5%w/v napthalene-1,5-disulphonic acid (NDS) [Eastman Kodak Company], 0.05M NaCl, 0.01M Tris HCl pH7.4) and left on ice for 1 min. to lyse. An equal volume of phenol mix (phenol [BDH] containing
14% v/v m-cresol [BDH], 0.1% w/v 8-hydroxyquinoline [BDH], saturated with 100mM Tris pH7.5) was added and the sample shaken for 20 seconds to mix the two phases. The lysate was centrifuged for 5 minutes at 3000xg to separate the phases and the upper aqueous layer removed for further purification, avoiding material at the interface. Phenol extraction was repeated until there was no sticky white interface (ca. 3 extractions) indicating removal of most of the protein contamination. A 1/10 volume of 3M NaAc pH5.6 and three volumes of ethanol and were added to precipitate total nucleic acid at -20°C overnight. Samples were centrifuged at 10000rpm for 30 min. to pellet the nucleic acid, the pellet was washed twice with 70% ethanol, recentrifuged and then dried by vacuum desiccation and resuspended in 50μl of sdH2O.

DNA was removed by treatment with DNAsase. An equal volume of DNAsase buffer (100mM 2-[N-morpholino]ethanesulphonic acid (MES) [SCC], 5mM magnesium acetate (MgAc) [BDH] pH7.0) was added to the sample and 1 unit of RNAse free DNAsase enzyme added per μg of DNA estimated to be present in the preparation. This mixture was left for 20' on ice before the reaction was terminated by the addition of 200mM NaCl. An equal volume of phenol mix was used in a single extraction before the sample was ethanol precipitated overnight at -20°C. The RNA was recovered by centrifugation, washed and dried as before and resuspended to a concentration of 2μg/ml in sdH2O.

2.2.7 Preparation of Total RNA from Sunflower

For some of the RNA transcript analysis described in Chapter 5, RNA from flower tissue was used. Attempts were made to purify mitochondria specifically from anther tissue, however the aborted nature of these structures in the sterile line made this impossible and hence a comparison with tissue from the fertile line infeasible. Instead, anthers and flowers at various stages of development were used for total RNA extraction.

Individual flowers or parts thereof were removed from sunflower inflorescences using microforceps and, for very immature tissue, a magnifying glass was required. Length of the flower buds was used as an indication of development stage in both fertile and sterile lines. The flowers were placed immediately into liquid nitrogen and kept frozen until use in RNA extraction. Initially, total RNA was made by grinding tissue in liquid nitrogen in a pestle and mortar then adding NDS buffer containing a drop of phenol mix, at approximately 2ml/g of tissue. The sample was purified by phenol extraction and precipitated in ethanol as for mtRNA. However, yields of total RNA produced in this way were variable and generally not high (2-5μg/g fresh tissue).
A second method, based on that of Castresana et al., (1988) reproducibly yielded approximately 10μg total RNA/g fresh fertile plant material. Tissue was ground to a fine powder in liquid nitrogen before the addition of guanidinium buffer (5M guanidinium thiocyanate [Fluka Biochemika], 25mM sodium citrate, 0.5%v/v sarcosyl, 2mM EDTA, 1mM β-mercaptoethanol [BDH] and 50mM Tris-HCl pH 7.6) at a ratio of 5ml/g tissue. The homogenate was transferred to a sterile Corex tube and shaken vigorously. This was centrifuged at 5000xg for 10 min. and the clarified supernatant extracted 2-3 times with an equal volume of phenol/chloroform (50%v/v phenol, 50%v/v chloroform saturated with 100mM Tris-Cl pH7.5) until there was no obvious protein material at the interface. Total nucleic acid was precipitated in 2.5 volumes ethanol and 0.3M NaAc at -20°C, overnight. After collection, the pellet was resuspended in 4ml of sdH₂O, lithium chloride (LiCl) [BDH] was added to a final concentration of 2M and the solution left on ice, overnight for RNA to precipitate. The insoluble RNA was pelleted by centrifugation at 10000xg for 10 min., washed with 70% ethanol, resuspended in sdH₂O and stored at -80°C. The supernatant from the LiCl precipitation was ethanol precipitated to recover total DNA suitable for genomic analysis (although this fraction was contaminated with tRNA and small RNA molecules).

2.2.8 RNA Gel Electrophoresis and Hybridisation

RNA was fractionated by electrophoresis on denaturing gels of 1.3%w/v agarose. This was made by melting 1.3g of agarose in 85ml of sdH₂O in a microwave, and allowing it to cool to 60°C before adding 10mls of MOPS buffer (0.2M MOPS, 50mM NaAc pH7.0, 10mM EDTA), 5.4mls of 37% formaldehyde [BDH] solution and EtBr to a final concentration of 0.2μg . This gave a 0.6M gel with respect to formaldehyde which was found to be adequate for denaturation and had a less deleterious effect on staining of the gel than the 2.2M formaldehyde recommended by some protocols. RNA samples (ca. 1-10μg in 1-10μl sdH₂O) were added to an equal volume of formamide sample buffer (50mM MOPS, 47%v/v formamide [BDH] 11%v/v formaldehyde), heated to 65°C for 2 min., quickly cooled on ice and then loaded onto the gel. Gels were run at room temperature at 150mV for 3 hours. No pretreatment of gels was necessary for subsequent transfer of the RNA onto Hybond-N. Traditional capillary blotting with 20xSSC gave more efficient transfer of RNA than the vacuum aided transfer described for mtDNA blotting . Hybond-N filters were UV treated for 2 min. and pre-hybridised at 42°C in 5xSSC, 50%v/v formamide, 5x Denhardtts, 0.1%v/v SDS and a final concentration of 400μg/ml sonicated herring sperm DNA. Following hybridisations, washes were as for DNA filters with respect to composition, but carried out at 42°C. For stripping probes from RNA blots, the filter was placed at 65°C for 1-2 hours in 5mM Tris-HCl pH8.0, 2mM EDTA and 0.1x Denhardtts then dried before repeated prehybridisation.
2.2.9 Preparation of chloroplast DNA

A rapid method for isolating chloroplast DNA (cpDNA) was used to prepare small quantities from green leaves of sunflower plants grown under light conditions. 30g of freshly harvested leaf tissue was homogenised for 15 seconds in 100mls of chloroplast grinding buffer (0.33M sorbitol, 50mM Tris-Cl pH8.4, 4mM β-mercaptoethanol, 1mM Na₂EDTA and 1mM MgCl₂). The homogenate was filtered through 4 layers of muslin into four 50ml corex tubes which were spun at 3000xg for 2 min. Supernatants were discarded and the pellets very gently resuspended with a piece of cotton wool, soaked in buffer and wound round a glass rod. The pellets were pooled in a total of 45mls of grinding buffer and then centrifuged as before. The pellet was resuspended in 10mls of chloroplast DNA extraction buffer (100mM Tris-HCl pH8.2, 6%w/v 4-aminosalicylate (PAS) [SCC] and 1%v/v triisopropyl napthalene sulphonate (TNS) [Kodak] and an equal volume of phenol-chloroform mix (see 2.2.7) was added. The sample was vortexed vigorously for 10 seconds then spun for 5 min. at 5000xg. The phenol extraction was repeated three times before ethanol precipitation of the DNA overnight. The DNA precipitate was recovered by centrifugation and resuspended in 10mls of TE8.0 to which 9.7g of cesium chloride was added. The sample was loaded into a heat-sealable tube and spun in the 70Ti rotor at 46000rpm for 20 hours. Chloroplast DNA, identified as a broad, faint band which fluoresced under UV-transillumination was removed from the cesium gradient and subsequently purified as described for mtDNA.

2.2.10 Cosmid Cloning of Sunflower Mitochondrial DNA

A library containing the entire mitochondrial genome from both fertile and sterile sunflower was constructed in the cosmid vector pJB8 (Ish-Horowicz and Burke, 1981). Cosmids were specifically designed for the cloning of large fragments of eukaryotic DNA and their use combines features of both bacteriophage λ and plasmid cloning (Maniatis et al. 1982). Cosmid vectors include the cos (cohesive end site) of λ at which unit lengths of DNA are circularised from longer concatemeric molecules and packaged into phage particles. Appropriate lengths of foreign DNA can be inserted between cos sites to produce hybrid molecules which mimic the natural packaging substrate in vitro. Once introduced into E.coli, such recombinant cosmid molecules can be selected on the basis of antibiotic resistance and will replicate in the form of large plasmids thus amplifying copies of the insert DNA. The pJB8 vector is a modified form of the Homer I ampicillin resistant (Amp<sup>5</sup>) cosmid which is derived from the plasmid pAT153 (Twigg and Sheratt 1980, Ish-Horowicz and Burke 1981). Mitochondrial DNA was cloned into the BamHI linker of the pJB8 vector by a procedure devised by Ish-Horowicz and Burke (1981) and outlined in Figure 2.1.
Procedure for the Cloning of High Molecular Weight Mitochondrial DNA Fragments into the Cosmid Vector pJB8

The cosmid pJB8 was separated into two aliquots for the creation of vector 'arms' by digestion with either HindIII or Sal I. These fragments were dephosphorylated with alkaline phosphatase before subsequent digestion with BamHI. The BamHI- HindIII and BamHI-SalI fragments containing the pJB8cos sites were isolated by DEAE purification (as described in methods) for ligation with partially digested mtDNA. The enzyme Sau3A was used to partially digest mtDNA from fertile and sterile sunflower under conditions to optimise formation of large genomic restriction fragments. Partial digested mtDNAs were fractionated on a cesium chloride density gradient and the fractions containing only fragments in the range 30-50kb were pooled. 250μg of size selected mtDNA was dephosphorylated with alkaline phosphatase prior to ligation with cosmid DNA fragments. Following ligation, *in vitro* packaging was used to package those products of ligation containing an mtDNA insert of 32-45kb between two cos sites within flanking pJB8 vector arms. Packaged pJB8 recombinant molecules were then transformed into *E.coli* and amplified by growth in ampicillin selective growth medium.
Figure 2.1 Cloning of mitochondrial DNA into the cosmid vector pJB8

(i) Preparation of vector

![Diagram of vector preparation]

(ii) Preparation of mitochondrial DNA

Native mtDNA

Partial digestion with Sau3A

Fractionation on a cesium chloride gradient and selection of fragments within the range 30-50kb

dephosphorylation of 30-50kb mtDNA fragments

(iii) in vitro packaging and infection of *E. coli*
i. Preparation of vector DNA. 5µg pJB8 cosmid DNA was digested to completion with HindIII or SalI and the reactions terminated by the addition of 200 mM EDTA. Linearised fragments were then dephosphorylated with 2 units of calf intestinal alkaline phosphatase (CIP) in CIP buffer (50mM Tris-Cl pH9, 1mM MgCl₂, 100nM zinc chloride, ZnCl₂) at 37°C for 30 min. The phosphatase treatment was checked by using 100ng of the digested fragments in a test ligation reaction with 1 unit of ligase in 9µl ligase buffer (50mM Tris-HCl pH7.6, 10mM MgCl₂, 10mM dithiothreitol (DTT) and 1mM ATP pH7.0). Gel electrophoresis of the test ligation confirmed that only linearised fragments were present, and that no higher molecular weight DNA bands due to multiple fragment ligation were in evidence (the phosphatase treatment is necessary to prevent the formation of tandem vectors during ligation). The vector DNAs were then digested with BamHI to create the vector 'arms' able to accept insert mtDNA fragments during subsequent ligation. The large BamHI-HindIII and the small BamHI-SalI pJB8 fragments (which both contain the cos sites) were isolated by DEAE purification (see 2.2.14) and ethanol precipitated before ligation with mtDNA.

ii. Preparation of insert DNA. MtDNA prepared from fertile and sterile sunflower was incubated with Sau3A under conditions deliberately designed to give only partial digestion with this enzyme. This enzyme cleaves at the four-base sequence 5' GATC 3' which will occur on average approximately every 250bp within the mtDNA. This means that an essentially random population of fragments is produced during the digestion reaction and reduces the chance of selective omission of specific regions of the mitochondrial genomes during the cloning procedure. The products of partial digestion were then separated on CsCl step density gradients (with density layers of 0.6mls each of ρ=1.7-1.44-1.39-1.34-1.29-1.24-1.19). The digests were layered carefully on the top of the gradients, which were then centrifuged at 40000xg for 90 min. at 20°C in an AH629 rotor. The gradients were fractionated by puncturing a hole in the base of the gradient tubes with a fine gauge needle and allowing 10-12 drop fractions to collect in a microtitre well (approximately 250µl per well). An aliquot of each fraction was analysed by gel electrophoresis and those fractions having a mobility equivalent to intact λ DNA (ca. 30-50kb) were pooled and dialysed against TE8.0 for 24 hours with two changes of the buffer before being ethanol precipitated and resuspended in TE8.0. 250µg of fractionated mtDNA were dephosphorylated at 37°C for 40 minutes in a 20µl reaction volume containing 1 unit of CIP enzyme in CIP buffer. After incubation the mtDNA samples were extracted once with phenol, ethanol precipitated and resuspended in TE8.0.

iii Ligation and Packaging. A volume of 20µl ligase buffer was used for ligation which contained equimolar ratios of size selected phosphatased mtDNA and the BamHI cleaved dephosphorylated HindIII or SalI pJB8
fragments (2.9μg mtDNA, 150ng each vector fragment) plus two units of T4 DNA ligase. The ligation mixes were left for 8 hours at 22°C before aliquots were analysed by agarose gel electrophoresis for evidence of high-molecular weight ligation products.

Approximately 0.5μg DNA in 3μl of ligation mix were used for *in vitro* packaging with a packaging kit according to the manufacturers instructions [Amersham International plc.]. After two hours at room temperature, 500μl of phage dilution buffer (10mMTris-Cl pH7.4, 10mM magnesium sulphate, MgSO₄, 0.01%w/v gelatin) and 10μl chloroform were added to give packaged phage solution. The packaged cosmids stored over chloroform in this way were stable at 4°C for several months. The only packageable molecules present in the ligation mix are those containing one of each of the vector fragments (and thus the two cos sites) separated by an insert fragment of between 32-45kb (see Figure 2.1) The dephosphorylation and size selection of the target DNA eliminates the the possibility of ligation of smaller, non-contiguous mtDNA fragments together to a packageable size of 'insert'. In addition, the vector fragments can neither self-ligate nor ligate in tandem thus preventing the formation of cosmids consisting of multiple vector inserts.

iv. Cosmid Transformation. The evening prior to cosmid transformation, a 100ml L-broth (LB)-maltose (LB containing 0.2%w/v maltose [SCC] where LB is 1%w/v tryptone [DF], 0.5%w/v yeast extract [DF], 0.5%w/v NaCl pH7.2) was inoculated with a single colony of *E.coli* ED8767 and left shaking overnight in a 37°C incubator. The following morning 1ml of the culture was used to inoculate 100ml of fresh LB-maltose and was grown for 4-5 hours until the optical density reading at 600nm was 1 (corresponds to ca. 8 x 10⁸ cells per ml). The bacteria were then concentrated 10-fold in 10mM MgSO₄ for use as phage sensitive indicator cells in the transformation procedure.

20μl of packaged phage solution was added to 100μl of indicator cells and incubated at 37°C to allow for phage adsorption. This was done in duplicate, one mix was undiluted and the other was serially diluted before plating on 7cm² bioassay plates of L-agar (LB containing 1.5%w/v agar) containing 100μg/ml ampicillin. The plates were incubated overnight in an inverted position at 37°C. 10μl of packaging solution which had not been added with any ligation mix was also transformed to confirm that background contamination had not contributed to the colonies produced. The following day, a plate showing easily identifiable individual colonies was used to calculate the transformation efficiency. Single colonies were picked onto ordered grid petri-dishes using sterile toothpicks so that each colony had a reference position for subsequent screening and the entire collection became a library. After overnight incubation at 37°C, the libraries were duplicated into 96-well microtitre plates containing 250μl per well of freezing broth-ampicillin.
(100μg/ml ampicillin in 0.5% yeast extract, 1% tryptone, 0.5% NaCl, 0.63% K₂HPO₄, 0.045% sodium citrate, 0.009% MgSO₄, 0.09% (NH₄)₂SO₄, 0.18% KH₂PO₄ all w/v and 4.4% v/v glycerol [BDH]) and grown up overnight at 37°C.

These duplicate frozen library stocks were kept at -80°C while working stock agar plates were duplicated regularly and stored at 4°C. For screening purposes, libraries were duplicated by direct transfer onto nitrocellulose filters which were then treated so that most of the host bacterial chromosomal DNA was removed and cosmid DNA retained. DNA within the filter bound colonies was denatured (5 min. on an absorbent pad of 3MM paper soaked in 1.5M NaCl, 0.5M NaOH), neutralized (5 min. on 1.5M NaCl, 10mM EDTA, 0.5M Tris- HCl pH7.2), rinsed in 5xSSC and fixed by baking in a vacuum oven at 80°C for 2 hours. Replica filters of cosmid clones were hybridised to radioactive probes using hybridisation conditions described in 2.2.5. Clones giving signals above background were selected for further analysis by small scale DNA preparation (miniprep) and dot blot hybridisation. Positives from this stage were used for large scale DNA preparation and restriction endonuclease analysis.

2.2.11 M13 Cloning

M13 is a single-stranded phage modified for use as a screenable cloning vector by the introduction of part of the β-galactosidase gene from E.coli (Messing et al. 1977). This lac region complements the lesion in certain lac-c hosts (e.g. JM101). The M13 vectors mp18 and mp19 have ‘polylinker’ regions containing sites for several restriction endonuclease enzymes within the lac region. Cloning into the polylinker insertionally inactivates the lac locus (Norrander et al. 1983). In the presence of an inducer of the lac operon Isopropyl-β-D-thiogalactoside (IPTG) and a chromogenic substance 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), host JM101 cells which have been transformed with M13 DNA form blue plaques when plated on a lawn of untransformed bacteria, while cells transformed with recombinant M13 produce white plaques.

This system was used to generate single-stranded M13 clones of the ATP gene and flanking regions for sequence analysis (Chapter four). Inserts from plasmid clones were purified by electroelution onto DEAE paper (see 2.2.14) and digested with Sau3A for ligation into the BamHI site of mp18 and mp19 or with TaqI for ligation into the AccI sites of mp18 and mp19. Vector DNA was digested and DEAE purified to ensure that only the linear form was present in the ligation. Ligations were performed in a total volume of 10μl containing approximately 60ng mtDNA and 20ng vector plus 1unit T4 ligase for 4 hours at room temperature.
Overnight cultures of JM101 were diluted 100 fold in 100ml LB and grown for 90 min at 37°C. Cells were pelleted and resuspended twice in 10ml of cold 100mM CaCl₂, then gently resuspended in 2mls and left on ice for at least 1 hour to render them competent for transformation. 200µl of competent cells were used per transformation with 5µl of the ligation mix and left on ice for 40 min. The cells were heat-shocked by incubating at 37°C for 5 min. then returned to ice. 200µl of log phase JM101 ('plating cells') , 600µg X-gal [NBL enzymes Ltd] and 480µg IPTG [NBL enzymes Ltd] for colour selection and 2ml of molten BBL top (1% w/v agar, 0.5% w/v NaCl) were added, before plating the transformation mix on minimal plates (1.5% agar, 0.2% (NH₄)SO₄, 1.4% KH₂PO₄, 0.6% K₂HPO₄, 0.1% sodium citrate, 0.02% MgSO₄, 0.2% glucose [BDH] all w/v)

2.2.12 Plasmid Cloning

Restriction endonuclease fragments from cosmids were subcloned into the plasmids pUC 18 or pAT153. These vectors contain the lac region of M13 allowing the white-blue colour screening procedure described above to be used. Transformed cells were selected by growth in the presence of ampicillin. EcoRI fragments containing the atpA genes and BamHI fragments containing the cob genes of fertile and sterile sunflower were cloned into pUC18 for restriction endonuclease mapping. A 1.2kb EcoRI fragment unique to sterile sunflower was cloned into the vector pAT153. Procedures for picking, and screening of plasmid colonies were as described for cosmids.

2.2.13 Preparation of Cloned DNA

i. Preparation of M13 single-stranded DNA. Recombinant M13 single-stranded DNA was prepared from phage culture as described by Messing et al. (1981) except that the growth medium was LB. Phage were picked from white plaques into 1ml log-phase JM101 and grown at 37°C for 4 hours. Cultures were then centrifuged in a microfuge for 3 min and phage particles were precipitated from the supernatant by the addition of 200µl of 20% w/v type 6000 polyethylene glycol (PEG) [BDH], 2.5M NaCl for 30 min. at room temperature then centrifuged again for 5 min. After complete removal of the PEG supernatant solution, phage DNA was purified by two phenol extractions and ethanol precipitation before final resuspension in 50µl TE.

ii. Preparation of plasmid DNA. Plasmid DNA was purified by a modification of the alkaline lysis procedure of Birnboim and Doly (1979). Two scales of preparations were used, termed mini and maxi. For mini-scale preparations, cells picked from a single colony were grown overnight at 37°C in 5mls of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl pH7.2, all w/v) containing 100µg/ml ampicillin. Cells from 2ml of overnight culture were pelleted in a microfuge for 30 seconds then resuspended in 100µl of lysis buffer
(25mM Tris-HCl pH8.0, 10mM EDTA, 1%w/v glucose, 2mg/ml lysozyme) at room temperature. After 5', 200μl of cold 0.2M NaOH/1%v/v SDS was added, gently mixed and the mixture placed on ice. After 5', 150μl of cold 3M NaAc, pH 5.0 was added. Bacterial chromosomal DNA was pelleted out after 5' on ice, leaving plasmid DNA in the supernatant which was deproteinised by phenol extraction. Plasmid DNA was precipitated in ethanol and finally resuspended in 50μl TE. Maxi-scale plasmid DNA preparations used the pellet from a 500ml overnight culture treated as for the minipreps but with all volumes scaled up by a factor of 100. The pellet after ethanol precipitation was resuspended in 2ml TE to which 1.9g CsCl and 125μl EtBr (5mg/ml) were added. This DNA solution was placed in 3ml capacity heat-seal tubes for centrifugation in a TL100 ultracentrifuge at 80,000rpm, overnight at 12°C. Once banded in CsCl, further purification was as described in 2.2.3.

iii. Cosmid DNA preparation. Two scales of preparation were also used for the isolation of cosmid DNA. The mini-scale described used 10ml overnight culture, and the maxi used the same protocol scaled up by a factor of 20. Selected clones were grown overnight in LB-ampicillin and pelleted in 15ml polypropylene tubes by centrifugation at 3000g for 10 minutes. Pellets were resuspended in 50μl ST buffer (sucrose 25%w/v, 50mMTris-HCl pH8.0) and transferred to a 1.5ml microfuge tube. 25μl of 20mg/ml lysozyme in ST buffer was added and incubated at room temperature for 15 min. before 63ml 200mM EDTA pH8.0 was added and left on ice, for a further 15 min. Then, 163ml TL buffer (Triton X100, 1%v/v, 62.5mM EDTA, 50mM Tris-HCl pH8.0) was added and mixed gently and kept on ice for 15 min. The lysate was frozen at -80°C then thawed at 37°C to encourage cell breakage. Bacterial debris was pelleted by centrifugation for 5 min. in a microfuge and removed with a pasteur pipette. The supernatant was phenol extracted, ethanol precipitated, resuspended in TE and used without further purification in the case of mini-preparations. Large scale preparations of cosmid DNA required CsCl purification as described for plasmid DNA.

2.2.14 DEAE Gel Purification

Restriction endonuclease digests of plasmid clones were separated by gel electrophoresis in agarose of varying percentages (0.6-1%w/v) depending on the size of fragment to be purified. Strips of DEAE membrane (prewetted in TBE) were inserted into the gel, in cuts made either side of the fragment using a clean scalpel. By resuming electrophoresis the fragment migrated onto the membrane placed in front of it and was protected from contamination by the larger fragment by the membrane behind it. The DNA of the collected fragment was then eluted from the membrane by a 20 min. incubation, at 65°C, in 200μl of DEAE elution buffer (1.5M NaCl in TE). The solution was phenol extracted, ethanol precipitated, dried, resuspended in sdH2O and
recovery checked by electrophoresis of a small aliquot. Double-stranded fragments purified in this manner were used for M13 cloning and random-primer labeling.

2.2.15 Preparation of Radioactive Probes

DNA clones for use in hybridisation to DNA and RNA blots were labelled using α^{32}P-dCTP by several methods. In all cases trichloroacetic acid (TCA) precipitable counts were used to estimate incorporation levels and unincorporated nucleotides were removed by purification on a 0.8ml column of Sephadex G50 [Pharmacia] equilibrated in TE according to Maniatis et al. (1982). Probes were boiled for 2 minutes and quickly cooled on ice to ensure that they were single-stranded before being added to the hybridisation buffer.

i. TCA precipitation. A solution of 5% w/v TCA [BDH] was placed in two petri dishes, and ethanol into a third. Two µl of the labeling reaction were removed, spotted onto a GF/C filter and dried with a hair drier. The radioactive counts on the filter were estimated by Cerenkov counting. Then the filter was placed in one of the dishes of TCA and agitated for 5 minutes then into the second TCA dish for a further 5 minutes. The filter was then rinsed briefly in ethanol, dried and re-counted. The percentage incorporation of label was calculated as the second count x 100/first count.

ii. Nick Translation. The method of 'nick translation' (Rigby et al. 1977), was used to label double stranded cloned DNA. The level of incorporation for this method ranged from 40-70%, giving probes with a specific activity up to 10^{7}dpm/ug DNA.

iii. Primer Extension. Single-stranded M13 clones were labeled by second strand-synthesis by a modification of the method of Hu and Messing (1982). A oligodeoxynucleotide sequence homologous to the polylinker of M13 was annealed to the template and used to prime second strand synthesis by the Klenow fragment of DNA polymerase. 0.2µg of DNA with 5ng primer in annealing buffer (10mM Tris-HC1 pH8.0, 100mM MgCl2) were heated at 65ºC for 20 min. The mixture was then allowed to cool to room temperature for the annealing to complete. 2mM of each of dATP, dGTP and dTTP [all BM] were added, plus 20µCi of α^{32}PdCTP and 6 units of Klenow. After 2 hours at 37ºC, 100µl TE8.0 was added to stop the reaction and the labeled DNA purified from unincorporated as before. Incorporation levels were usually >60% and probes had a specific activity in the range 10^{8}dpm/µg DNA.

iv. Random Probe Priming. Restriction endonuclease fragments purified by DEAE membrane elution (2.2.14) were labelled by oligonucleotide random priming (Fernberg and Vogelstein 1984, Hodgson and Fisk 1987). 50ng insert DNA plus 20ng random hexanucleotide primers were heated in TE8.0 to 100ºC for 2 min. The mixture was then placed on ice, 2µl of 10x labelling buffer was added (500mM Tris-HCl pH7.0, 100mM MgSO4, 1mM DTT and 6mM each of dATP, dGTP and dTTP) plus 30µCi of α^{32}PdCTP, 6units Klenow and sdH2O to
bring the total volume to 20μl. The reaction was left at room temperature for
3 hours and terminated by the addition of 100μl TE8.0 before purification on
G50 Sephadex. Incorporation levels varied but were normally greater than
80% and probes with specific activities in excess of 10^9 dpm/μg DNA were
obtained.

v. Oligonucleotide End Labelling. An oligonucleotide, complementary to
the \textit{nad3} gene from maize, was end-labeled using γ-32PdATP. 75ng (10pmol) of
the oligonucleotide was incubated with 30μCi (10pmol) γ-32PdATP and 10
units of T4 polynucleotide kinase in kinase buffer (0.5m Tris-HCl pH7.6, 0.1M
MgCl₂, 50mMDTT, 1mMEDTA) at 37°C. After 1 hour incorporation was
calculated to be >90% and the entire reaction mixture was added to
hybridisation buffer without further purification. The required temperature
for annealing of the oligonucleotide to sunflower mtDNA was calculated
according to the formula of Wallace \textit{et al.} (1979) and the base composition of
the oligonucleotide to be 63°C.

vi. Mitochondrial DNA End-Labelling. MtDNA was end-labelled with
γ-32PdATP by a method based on that of Maizel \textit{et al.} (1976). 2μg of sunflower
mtDNA was heated to 90°C for 5 minutes in alkaline hydrolysis buffer (50mM
glycine pH to 9.5 with NaOH) in a reaction volume of 10μl. After cooling to
room temperature, 2μl of the sample was added to a labelling mix containing
70mM TrisHCl pH7.6, 10mM MgCl₂, 5mMDTT, 10μg BSA, 5 units of T4
polynucleotide kinase and 40μCi of γ32PdATP. The total reaction volume was
made up to 40μl with sdH₂O and the mix was incubated at 37°C for 1 hour.
After incubation, the sample was ethanol precipitated three times and finally
dried and resuspended in 100μl sdH₂O and added to the hybridisation mix.

2.2.16 DNA Sequencing

M13 DNA clones were sequenced using a DNA Sequenase Sequencing
Kit supplied by United States Biochemical Company and [α-35S]dATP. The
manufacturers protocol for the chain termination method for sequencing was
used (Sanger \textit{et al.} 1977). The Sequenase enzyme can be used to optimise
production of long DNA chains allowing sequences which extend up to 500
bases from the priming site to be read in a single experiment. For
electrophoretic separation of the reaction products of sequencing reactions
0.4mm thick 8% acrylamide gels (8M urea [BDH], 8%w/v acrylamide [Eastman
Kodak Company], 0.25%w/v N,N′-Methylenebisacrylamide (bisacrylamide)
[BDH] in 1xTBE) containing 0.1% w/v ammonium persulphate (APS) [BDH]
and 0.005% N,N,N′,N′-Tetramethylethylenediamine (TEMED) [SCC]) were used.
Gels were electrophoresed for 2-7 hours to resolve shorter or longer chains as
required. After electrophoresis, gels were soaked in 10%v/v methanol [BDH],
5%v/v glacial acetic acid [BDH] for 10 min. then rinsed in dH₂O before being
dried, under vacuum, onto 3MM chromatography paper for autoradiography.
2.2.17 Mitochondrial Protein Synthesis

i. Incorporation of $^{35}$S-methionine into isolated mitochondria

Mitochondria prepared by differential centrifugation and sucrose purification were used for incorporation of radiolabeled L-$^{35}$S-methionine into proteins synthesised \textit{in organello} by a method based on that of Leaver et al. (1983) with modifications by A.D. Liddell (personal communication). For each mitochondrial incorporation three alternative energy sources were used:

1) \textbf{Acetate} (NaAc, 20mM) - a non-oxidisable substrate that cannot be used by mitochondria and which gives an indication of the level of bacterial contamination; 2) \textbf{Creatine phosphate /creatine kinase} (CP/CPK, 8mM/25µg) - provides externally generated ATP to the mitochondria; 3) \textbf{Succinate/ADP} (succ/ADP, 10mM/1mM) - mitochondria use succinate as a substrate from which to generate their own ATP by oxidative phosphorylation.

The incubation mixture used for incorporation consisted of 1.6mM guanosine triphosphate (GTP), 3mM DTT, 300mM mannitol, 150mM KCl, 16mM tricine pH 7.2, 8mM KPO$_4$ pH 7.2, 16mM MgCl$_2$, 1mM EGTA pH 7.2, the 19 essential amino acids (approximately 3µM with respect to each) excluding methionine and 0.74MBq L-$^{35}$S-methionine (20µCi). One of the above energy mixes was added to the incubation mixture plus 50µl of a mitochondrial suspension (5-15mgs protein/ml in resuspending buffer) and incubated at 25°C with vigorous shaking for 90 min. After incubation, small aliquots were removed for estimating incorporation by TCA precipitation and the remainder of the 'incorporated' mitochondria was pelleted by centrifugation in 1ml of cold resuspending buffer in a microfuge at 4°C for 10' and then quick frozen at -80°C until further use.

ii. Protein gel electrophoresis Mitochondrial proteins labelled during incorporations were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) so that individual polypeptides synthesised during the labelling period could then be detected by autoradiography. Mitochondrial pellets (from i.) were thawed and resuspended to 5-10mgs/ml in sample loading buffer (120mM Tris-HCl pH8.0, 10%v/v glycerol, 10 mM DTT, 2%v/v SDS, 0.01%v/v bromophenol blue [Aldrich Chemical Company]), heated to 80°C for 2 min. and cooled on ice for 1 minute. The samples were loaded with a microsyringe onto a 1.5mm gradient resolving gel made using equal volumes of 12% acrylamide (8mls of 30%w/v acrylamide:0.8%w/v bisacrylamide, 80µl 10%w/v APS, 10µl TEMED in 187.5M Tris-HCl, pH8.5) and 20% acrylamide (13.4mls 30%w/v acrylamide:0.8%w/v bisacrylamide, 80µl 10%w/v APS, 5µl TEMED containing 4.4g sucrose). The two solutions (20mls of each) were put into separate chambers of a gradient former placed on a magnetic stirrer, the dense (20%) solution was in the mixing chamber which contained a magnetic flea. The dense solution was allowed to fill the rubber tubing used to connect the
chambers to the gel plate assembly. A sealing screw between the solution chambers was then released so the light solution entered the mixing chamber and was thoroughly mixed with the dense solution. After both chambers were empty (approximately 15 minutes) a small amount of overlay buffer (resolving buffer containing 0.1%v/v SDS and 80%v/v isopropanol) was run onto the gel surface and the gel allowed to polymerise for 3-4 hours. The overlay was poured off before a stacking gel of 5% acrylamide (30% acrylamide: 0.2% bisacrylamide stock) in stacking gel buffer (60mM Tris HCl pH6.8) was added, the sample comb was inserted and the gel allowed to polymerise for 30 min. Approximately, 200μg total mitochondrial protein was loaded per track and molecular weight marker proteins loaded in the outer lanes were BSA (66kDa), catalase (57.5kDa), aldolase (39kDa), carbonic anhydrase (29kDa), trypsin inhibitor (20kDa), myoglobin (17kDa) and cytochrome c (12.5kDa). Gels were electrophoresed for 16 hours at 5mA in Laemmli electrode buffer (50mM Tris, 192mM glycine [SCC]) containing 0.1%v/v SDS.

After electrophoresis, the resolving gel was stained for 4-6 hours in 45%v/v methanol, 8%v/v glacial acetic acid, 0.2% w/v Coomassie blue [BDH] then destained in the same solution without dye, overnight. The gel was washed in 50% methanol, dried onto 3MM chromatography paper and then expose to X-ray film at -80°C for 1-3 weeks.
CHAPTER THREE

MITOCHONDRIAL GENES, GENOMES AND MOLECULAR REARRANGEMENT IN SUNFLOWER

3.1 Introduction and Rationale

In order to investigate mitochondrial genome variation associated with CMS, cloned gene sequences can be used as markers in DNA hybridisation experiments to identify points or regions of divergence between the genomes from cytoplasms which confer the fertile or sterile phenotypes. An inherent assumption in this approach, is that the mutation(s) causing CMS occurs in a sequence which has been previously identified as mitochondrially encoded and expressed and is therefore available as a probe.

As discussed earlier, the genes encoding several proteins which are synthesized in higher plant mitochondria have still to be identified and it is possible that the CMS determinant in sunflower mitochondria resides in some uncharacterised coding region. However, the demonstration in maize, sorghum and Petunia that the putative coding regions responsible for the CMS phenotype have been created by recombination within or near several known genes, strengthens the case for an investigation of genome reorganisation around known mitochondrial genes in sunflower.

3.2 Isolation of Native Mitochondrial DNA from Fertile and Sterile Sunflower

Mitochondrial DNA was isolated from gradient purified mitochondria as described in methods and the concentration estimated by measuring the absorbance of an aliquot at 260nm in a spectrophotometer. The estimated concentrations were checked by agarose gel electrophoresis of native mtDNA compared to a known concentration of undigested λ DNA as shown in Figure 3.1. Intact high molecular weight mtDNA migrates as a tight band at the level of the λ DNA (molecular weight of ca. 50kb). The yields of mtDNA obtained from both the fertile and sterile sunflower were of the order of 0.3-0.5μg/g fresh weight of etiolated hypocotyl and cotyledon tissue.

A low molecular weight band separate from the main high molecular band is present in mtDNA from fertile sunflower (Figure 3.1, lane 3). This is the 1.45kb circular plasmid P₁, previously described by Leroy et al. (1985 and section 1.6). The copy number of P₁ shown here for the 842B fertile line, is similar to that reported for the fertile line HA89B (also known as 841B) estimated to be present at 0.5-1 copy per mitochondrial genome (Crouzillat et al. 1987).
Figure 3.1
Comparison of native mitochondrial DNA from fertile and sterile sunflower.

Undigested mtDNA
Mitochondrial DNA (mtDNA) from sunflower etiolated seedling tissue was isolated by cesium chloride gradient purification as described in the text. 3μg of mtDNA from sterile sunflower (lane 2) or fertile sunflower (lane 3) were electrophoresed on 0.7% (w/v) agarose and compared with undigested lambda DNA (~50kb, lane 1). Ethidium bromide staining of the gel allowed visualisation of the mtDNA fragments when viewed by UV illumination. Note the presence of the low molecular weight band in mtDNA from the fertile cytoplasm of sunflower, this is the 1.45kb P1 plasmid molecule (arrow).

Partial digestion of mtDNA suitable for cosmid cloning
Conditions for partial digestion suitable for generating 30-40kb genomic fragments were determined by electrophoretic analysis of aliquots from digestions at several time points after the beginning of the reaction (see lanes 4-7).
Reduced levels of P₁ are present in some sunflower CMS cytoplasms although it is absent from others, including the sterile line (842A) shown in Figure 3.1 (lane 2). During the course of this investigation another plasmid-like DNA of ca. 5kb in size was identified in sunflower containing the restorer cytoplasm. The copy number of this plasmid was also found to vary with cytoplasmic type as discussed later in this chapter.

3.3 Estimation of Mitochondrial Genome Size by Digestion with Restriction Endonucleases

The restriction endonuclease patterns produced when mtDNA from fertile and sterile sunflower was digested with EcoRI or BamHI are shown in Figures 3.2 and 3.3 respectively. For both enzymes, the restriction patterns are similar for the two cytoplasmic types suggesting a lack of primary sequence divergence between them. However, some variation is present as indicated by a 12.4kb BamHI fragment in the sterile genome which is absent from the fertile genome (arrow in Figure 3.3). The variable intensity of many of the restriction fragments is due, in part, to unresolved doublet and triplet bands but also reflects the organisational complexity of the genomes.

By summation of the estimated sizes of all mtDNA fragments in a single 'restriction profile' estimates for the total size of the mitochondrial genomes from fertile and from sterile sunflower were calculated (Table 3.1). The values of 280-298kb are minimum estimates of the genome sizes due to the difficulty of resolving very small mtDNA fragments which may represent a sizeable portion of the genomes. Densitometer scans of photographs of gels were used to assess the stoichiometry of several bands. For example, the second largest EcoRI fragment of 9.2kb (Figure 3.2) was estimated to have a stoichiometry of two as the densitometer peak corresponding to this band was approximately twice that of the fragment immediately below. In general, the values obtained are in good agreement with the 305kb size calculated by physical mapping of the mitochondrial genome of the HA89A sterile line of sunflower by Siculella and Palmer (1988).
Figures 3.2 and 3.3
A Comparison of Mitochondrial DNA Restriction Profiles from Fertile and Sterile Sunflower Reveals Little Divergence

Figure 3.2 3μg of mitochondrial DNA (mtDNA) purified from fertile (lane 1) or sterile (lane 2) was digested to completion with EcoRI. The products were separated by electrophoresis through 0.8% w/v agarose at 25mA for 16 hours. Restriction fragments were visualised by ethidium bromide staining and sizes calculated by comparison to lambda (λ)—HindIII markers (sizes in kilobases, kb, shown between Figures 3.2 and 3.3). A densitometer scan of the gel photograph was used to estimate the relative stoichiometry of differently sized fragments for estimations of total genome size.

Figure 3.3 3μg of mtDNA from fertile (lane 1) or sterile (lane 2) sunflower was digested to completion with BamHI and electrophoresed as described above. The presence of a 12.4kb fragment in mtDNA from sterile sunflower which is absent from fertile is indicated by the arrow.

Table 3.1
Estimates of mitochondrial genome sizes for fertile and sterile sunflower by restriction fragment summation
Gels similar to those shown in Figures 3.2 and 3.3 were scanned densitometrically to estimate restriction fragment stoichiometry. The sizes of all detectable fragments from an individual digestion were summed and compared between the two genomes. Stoichiometric estimation for fragments less than 0.5kb was inaccurate and not included in the final summation. The estimate for sterile sunflower, with BamHI is larger due to the detectable presence of the specific high molecular fragment (12.4kb). Variations in stoichiometry of lower molecular weight fragments give the size differences noted with EcoRI and Xhol.
Figure 3.2

Figure 3.3

Table 3.1

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Summation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
</tr>
<tr>
<td>EcoRI</td>
<td>280</td>
</tr>
<tr>
<td>BamHI</td>
<td>285</td>
</tr>
<tr>
<td>Hind III</td>
<td>292</td>
</tr>
<tr>
<td>XhoI</td>
<td>287</td>
</tr>
</tbody>
</table>
3.4 Investigation of the Mitochondrial Genome Organisation in Fertile and Sterile Sunflower by Mt DNA Hybridisation Analysis

Sunflower mtDNA was digested with a variety of restriction enzymes and the products were separated on gels similar to those shown in Figures 3.2 and 3.3. These were blotted onto nylon membranes and hybridised with various mitochondrial gene probes from several other plant species (i.e. heterologous probe hybridisation) which had been labelled with $^{32}$P. The autoradiographs of these hybridisation experiments are shown in Figures 3.4.-3.17 and a summary of the mitochondrial gene homologies found in sterile and fertile sunflower mtDNA is presented in Table 3.2. A number of interesting points arose from the use of specific gene probes to characterise the mitochondrial genome organisation in sunflower and are discussed briefly at the end of this chapter. A more detailed discussion of those findings of specific interest with regard to CMS is presented in Chapter 7.

In general, the results of the mtDNA hybridisation experiments confirm the restriction endonuclease data that mitochondrial genome organisation is similar in fertile and sterile sunflower. Of the twelve protein or rRNA marker genes used as hybridisation probes, ten appear, in their immediate vicinity at least, to be unaltered between the two cytoplasms. Restriction fragment length polymorphisms (RFLPs) indicative of sequence divergence, were found with the maize gene probes for $F_1$ATPase $\alpha$-subunit ($atpA$ gene) and the apocytochrome subunit b ($cob$ gene).

In EcoRI digested mtDNA, the major $atpA$ homologous fragment of 2.2kb from fertile sunflower is replaced by a 3.8kb fragment in sterile sunflower (Figure 3.4). In HindIII digests, the main 5.3kb hybridising fragment in the fertile genome is replaced by a larger 9.3kb fragment in the sterile. The probe ALXR18, which covers 544 nucleotides of the 5' region of the maize $atpA$ gene identifies identical 9.7kb BamHI mtDNA fragments in fertile and sterile sunflower. In contrast, when BamHI digested mtDNA is hybridised with the $cob$ probe 1435 (covering the 5' coding and upstream region of the maize gene) a 10.1kb fragment in mtDNA from fertile sunflower is replaced by a 12.4kb fragment in sterile (Figure 3.6). When mtDNA is digested with EcoRI and HindIII this $cob$ probe does not discriminate between the cytoplasms.
Figure 3.4
Hybridisation of the maize atpA gene to mitochondrial DNA from fertile and sterile sunflower identifies genomic divergence
3μg of mtDNA from fertile (lanes 1, 3 and 5) and sterile (lanes 2, 4 and 6) sunflower digested with EcoRI (lanes 1-2) HindIII (3-4) or BamHI (5-6) was electrophoresed through 0.8%w/v agarose and transferred to Hybond-N nylon membrane. The membrane was hybridised with a 32P-labelled probe of the 5' region of the maize atpA gene (M13 clone ALXR18, this is an EcoRI-ClaI sequence covering the 544 nucleotides of the coding region, beginning 3 nucleotides downstream of the ATG start of the gene, Isaac et al. 1985a) HindIII digested λ DNA (sizes shown in kb) were used to calculate the sizes of hybridising sunflower mtDNA fragments. Altered hybridisation patterns in sterile sunflower occur with mtDNA digested with EcoRI and HindIII.

Figure 3.5
Subliminal homologies to the 3' region of maize atpA in Sunflower
3μg of mtDNA from EcoRI digested mtDNA fertile (lane 1), sterile (2), hybrid (3) and restorer (4) lines of sunflower were electrophoresed through 0.8%w/v agarose, transferred to Hybond-N nylon and hybridised to a 32P-labelled maize probe covering the 3' end of the atpA gene (M13 clone BLSC1, a SacI-ClaI sequence covering 451 nucleotides including the end of the gene and the 3' flanking region Isaac et al. 1985a). Faint hybridisation signals, homologous to the atpA 3' probe are indicated by arrows. Note the low level of the 2.2kb EcoRI, the major hybridising atpA fragment in the fertile genome, in the other three lines.

Figure 3.6
Different BamHI restriction fragments contain the cob gene in mitochondrial DNA from fertile and sterile sunflower
A duplicate filter to that shown in Figure 3.4 was probed with a 32P-labeled gene probe of the maize cob 5' coding region (M13 clone 1435). A 12.4kb hybridising fragment occurs in BamHI digested mtDNA from sterile compared to a 10.1kb fragment in fertile sunflower.

Figure 3.7
At least two regions of the sunflower mitochondrial genome are homologous to the atp9 gene of the maize
A duplicate filter to that shown Figure 3.4 was hybridised with a 32P-labeled probe of the maize atp9 gene. With EcoRI (lanes 1-2), HindIII (lanes 3-4) and BamHI (lanes 5-6), at least two fragments hybridise suggesting an atp9 gene duplication exists in the sunflower genome.
3.5 The Sunflower Mitochondrial Genome Contains Substoichiometric Homologies to the atpA Gene

In addition to the different fragments which hybridise in mtDNA from sterile and fertile sunflower, with the atpA 5' probe, when a probe covering the 451 nucleotides of the 3' end of the maize gene and its immediate flanking sequence is used (BLSC1), several other hybridising fragments, of variable intensities, were detected in both genomes and also in the hybrid and restorer lines (Figure 3.5, arrowheads at 8.5kb, 2.2kb, 2.0kb). These faint signals may be due to short stretches of homology to the probe which occur elsewhere in the genomes. If this were the case, the variation in the level of hybridisation of these additional bands would reflect the extent of homology which they contain in terms of sequence composition and/or length. However in the sterile, hybrid and restorer cytoplasts, in addition to the major 3.8kb EcoRI hybridising fragment which they all show, there is a low level of the 2.2kb EcoRI fragment homology which occurs at 'normal' stoichiometry in the mtDNA from fertile sunflower. It is unlikely that differences in the extent of hybridisation of the same mtDNA fragment within the genomes of the different cytoplasts is due to partial homology to the probe. A more likely explanation is that this, and possibly the other, faint hybridisation signals are due to substoichiometric molecules within the mitochondrial genome which contain alternative atpA arrangements. Very long exposure of autoradiographs of mtDNA from the fertile cytoplasm failed to detect the 3.8kb EcoRI fragment in this genome.

3.6 The atp9 gene May Be Duplicated in the Mitochondrial Genome of Sunflower

When sunflower mtDNA, digested with a range of restriction endonucleases, was hybridised with an atp9 coding region probe, in each case two mtDNA fragment were labelled with equal intensity. Figure 3.7 shows the result of atp9 hybridisation with EcoRI, HindII and BamHI digested mtDNA. This suggests that two copies of the gene or duplication of part of the gene are present within the mitochondrial genome of sunflower. The hybridisation for atp9 was identical in mtDNA from fertile and sterile sunflower thus, despite multiple regions of atp9 homology within the sunflower mtDNA there is no indication that any are differently organised in the genome from sterile compared to fertile.

When a probe of the atp6 gene from maize was hybridised with mtDNA from fertile and sterile sunflower, single restriction fragments in several different digests were identical in both cytoplasts. This indicates that a single copy of atp6 is present in both types of sunflower and is unaltered in its organisation between the two (Figure 3.12)
3.7 The *coxII* Gene in Sunflower Contains an Intron

An interesting result was the finding that sunflower mtDNA contains homology to an intron specific probe from the maize *coxII* gene as shown Figure 3.8. In mtDNA from both fertile and sterile sunflower digested with *HindIII*, a 3.6kb fragment which hybridises to the intron probe is the same fragment labelled by a probe for exon 1 of *coxII* gene (Figure 3.9). This indicates that the two sequences are contiguous within the sunflower mitochondrial genome and that the *coxII* gene does contain the intron sequence. The additional *EcoRI* (3.5kb) and *BamHI* (1.9kb) restriction fragments which hybridise to the intron probe are due to recognition sites for these enzymes occurring within the intron sequence. The individual major hybridising fragments for the *coxII* exon probe with sunflower mtDNA indicates that a single copy of the gene is present in the main mitochondrial genome. The genes for *coxI* and *coxIII* were similarly found to be single copy in sunflower mtDNA and both appear to be unaltered between the fertile and sterile cytoplasms (Figures 3.10 and 3.11 respectively).

3.8 Homology to Sequences Encoding Subunits 1, 3 and 5 of the NADH dehydrogenase are Present in Sunflower Mt DNA

The result shown in Figure 3.13 indicates that homology to exon 2 of the *nad1* gene from *Oenothera* is clearly present as a single-copy sequence, which is identical in mtDNA from fertile and sterile sunflower. Probes for the *nad1* intron or exon 1 sequences which occur in some fungal mitochondrial genomes were not available and so the extent of *nad1* homology in the sunflower genome could not be determined precisely. An oligonucleotide containing 20 nucleotides of the 3' coding region of the maize *nad3* gene was used as a probe against sunflower mtDNA. Homology in the sunflower genome was located on individual restriction fragments with both *EcoRI* and *BamHI* digested mtDNA indicating that single, identical copies of this gene exist in genomes of fertile and sterile sunflower (Figure 3.14).

When a probe containing the region encoding the *nad5* gene from *Oenothera* was hybridised with mtDNA from sunflower, the fragments which hybridised intensely in the fertile and sterile cytoplasms were of identical sizes (Figure 3.15). Thus, neither the subunit 1, 3 nor 5 NADH dehydrogenase sequence homologies in sunflower mtDNA show any indication that they are present in altered genomic environments in the two cytoplasmic types.
Figure 3.8 The sunflower coxII gene contains an intron
3μg mtDNA from fertile (lanes 1, 3 and 5) and sterile (2, 4, and 6) sunflower was digested with EcoRI (lanes 1 and 2), HindIII (lanes 3 and 4) or BamHI (lanes 5 and 6). A 32P-labelled probe internal to the intron of the coxII gene from maize hybridised to restriction fragments within the sunflower mitochondrial genome indicated homology is present in this genome.

Figure 3.9 Hybridisation of a coxII exon 1 probe to sunflower mtDNA
A filter identical to the one used in Figure 3.8 was hybridised to a maize probe specific to the first exon of the coxII gene. This 5’ region of the gene is contained on some of the restriction fragments homologous to the intron, indicating the intervening sequence is part of the coxII gene in sunflower and that it contains recognition sites for EcoRI and BamHI.

Figure 3.10 Identification of coxI homology in sunflower mtDNA
3μg mtDNA from fertile (lanes 1 and 3) and sterile (2 and 4) sunflower was digested with EcoRI (lanes 1 and 2), HindIII (lanes 3 and 4). A 32P-labelled probe of the coxI gene from maize identified, for each enzyme, identical single fragments in the two cytoplasms.

Figure 3.11 Identification of coxIII in sunflower mtDNA
A duplicate filter of Figure 3.10 hybridised with a 32P-labelled probe of the coxIII gene from Oenothera identified only equivalent sized hybridising fragments in mtDNA from fertile and sterile cytoplasms of sunflower.

Figure 3.12 A probe for the atp6 gene detects no divergence between sunflower cytoplasms
Hybridisation of a duplicate filter of Figure 3.10 with a 32P-labeled probe of the atp6 gene from maize indicated the genes to be located in similar genomic environments within the fertile and sterile cytoplasms.
Figure 3.13 A Single Copy Homology to \textit{nad1} in Sunflower mtDNA
4μg mtDNA from fertile (lanes 1 and 3) or sterile (lanes 2 and 4) were digested with \textit{EcoRI} (1 and 2) or \textit{HindIII} (3 and 4), fractionated by electrophoresis through 0.8% w/v agarose and transferred to Hybond-N. This filter was hybridised with a probe containing the 3' region of the \textit{nad1} gene from \textit{Oenothera}. Identical, single regions of \textit{nad1} homology are present in the mtDNA fertile and sterile sunflower.

Figure 3.14 Homology to the \textit{nad3} Gene exists in Sunflower MtDNA
A chemically synthesised oligonucleotide containing a 20 nucleotide sequence homologous to the \textit{nad3} gene from maize was hybridised to mtDNA from fertile (lanes 1 and 3) and sterile (lanes 2 and 4) sunflower which had been digested with \textit{EcoRI} (1 and 2) or \textit{BamHI} (3 and 4), fractionated by electrophoresis through 0.8% w/v agarose and transferred to Hybond-N. With each enzyme, identical fragments are labelled in the two cytoplasms.

Figure 3.15 The \textit{nad5} gene is mitochondrially encoded in sunflower
3μg of mtDNA from fertile (lanes 1 and 3) and sterile (2 and 4) sunflower were digested with \textit{EcoRI} (1 and 2) and \textit{BamHI} (3 and 4), electrophoresed and transferred as described above. Using a 32P-labelled \textit{nad5} probe from the \textit{Oenothera} mitochondrial genome encoding \textit{nad5}, identical hybridisation patterns were seen for the two cytoplasms. The faint hybridising restriction fragments are due to partial homology of the probe to a section of the \textit{rrn5} gene encoding 5S rRNA.

Figure 3.16 Homology to the \textit{rps12} gene is detectable in sunflower mtDNA
The filter used for probing of \textit{nad1} shown in Figure 3.13 was partially stripped of this probe before rehybridisation with an M13 clone containing an internal portion of the \textit{rps12} gene from wheat. Residual hybridisation signals for \textit{nad1} are apparent however, new weakly hybridising fragments due to the \textit{rps12} probe are seen (arrows).

Figure 3.17 The genes for \textit{rrn18} and \textit{rrn5} are linked in sunflower mtDNA
The filter shown in Figure 3.14 was stripped of the \textit{nad3} probe and reprobed with an \textit{rrn18} gene probe encoding the 18S rRNA from \textit{Oenothera}. The \textit{rrn18} homologous fragments with \textit{EcoRI} (lanes 1-2) and \textit{BamHI} (3-4) for fertile (1, 3) and sterile (2, 4) sunflower mtDNA are the same size as those identified by the \textit{Oenothera} \textit{rrn5} gene probe (see Figure 3.15) indicating that these two genes are linked in the sunflower mitochondrial genome.
Figure 3.13: nad1

Figure 3.14: nad3

Figure 3.15: nad5

Figure 3.16: rps12

Figure 3.17: rrm18
3.9 Sunflower Mt DNA Contains Linked 18S and 5S Ribosomal Genes

The probe containing the nad5 gene described above also contained a small region of the gene encoding the 5S mitochondrial rRNA (rrn5) from *Oenothera*. The rrn5 homology in sunflower mtDNA corresponds to the more faintly hybridising fragments in Figure 3.15. The sizes of the restriction fragments which hybridise to rrn5, correspond to those which hybridise when a probe for the rrn18 gene was hybridised with sunflower mtDNA (Figure 3.17). This suggests, that these ribosomal RNA genes are linked in the sunflower mitochondrial genome as found in all other higher plant mtDNAs. The region encoding these ribosomal genes gives the same restriction pattern in fertile and sterile sunflower.

3.10 The *rps12* gene is Encoded by the Sunflower Mitochondrial Genome

An M13 probe containing part of the *rps12* coding sequence from wheat was hybridised with mtDNA from sterile and fertile sunflower. Using a hybridisation temperature of 65°C and a salt concentration of 5xSSC only a weak labelling of sunflower mtDNA restriction fragments was found which may have been due to a poor specific activity of the probe. However, a stronger signal was obtained when the stringency was lowered to 55°C and 6xSSC (shown in Figure 3.16) indicating that a low level of homology between the genes could be responsible for the weak hybridisation. Homology to *rps12* was not reported on the published map of the sunflower genome (Siculella and Palmer 1988).

3.11 The Use of Cosmid Clones to Investigate Rearrangement in the Mitochondrial Genomes of Fertile and Sterile Sunflower

As mtDNA hybridisation analysis indicated that regions of the mitochondrial genome encoding *atpA* and *cob* show evidence of sequence divergence between fertile and sterile sunflower these genes were isolated from both mitochondrial genomes for further analysis. Cosmid clones containing long sections of mtDNA were required to investigate the genomic organisation around the *atpA* and *cob* genes in the two cytoplasts. Individual restriction fragments within the cosmids were subcloned into plasmid vectors for more detailed analysis of the coding regions and flanking regions of the two genes.
Table 3.2
Restriction Fragment Homologies to Selected Genes in Sunflower Mitochondrial DNA

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>EcoRI</th>
<th>HindIII</th>
<th>BamHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoplasm</td>
<td>S</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>Gene Probe</td>
<td>(kb)</td>
<td>(kb)</td>
<td>(kb)</td>
</tr>
<tr>
<td>atpA†</td>
<td>3.8</td>
<td>2.2</td>
<td>9.3</td>
</tr>
<tr>
<td>atp6</td>
<td>4.7</td>
<td>4.7</td>
<td>1.6</td>
</tr>
<tr>
<td>atp9</td>
<td>15.1,5.1</td>
<td>15.1,5.1</td>
<td>5.8,3.2</td>
</tr>
<tr>
<td>cob</td>
<td>3.6</td>
<td>3.6</td>
<td>9.3,3.7</td>
</tr>
<tr>
<td>coxI</td>
<td>4.9</td>
<td>4.9</td>
<td>4.6</td>
</tr>
<tr>
<td>coxII (intron)*</td>
<td>3.5</td>
<td>3.5</td>
<td>3.6</td>
</tr>
<tr>
<td>coxII (exon)</td>
<td>2.4</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td>coxIII</td>
<td>3.1</td>
<td>3.1</td>
<td>7.1</td>
</tr>
<tr>
<td>nad1</td>
<td>4.3</td>
<td>4.3</td>
<td>5.1</td>
</tr>
<tr>
<td>nad3</td>
<td>1.1</td>
<td>1.1</td>
<td>nd</td>
</tr>
<tr>
<td>nad5</td>
<td>1.9</td>
<td>1.9</td>
<td>nd</td>
</tr>
<tr>
<td>rps12</td>
<td>0.5</td>
<td>0.5</td>
<td>4.8</td>
</tr>
<tr>
<td>rrn18</td>
<td>2.3</td>
<td>2.3</td>
<td>nd</td>
</tr>
</tbody>
</table>

The estimated sizes (in kb) of restriction fragments of mtDNA from fertile (F) and sterile (S) sunflower which show hybridisation with specific mitochondrial gene probes are shown. † Denotes the main atpA homologous restriction fragments noted using a probe specific to the 5’ coding region (ALXR18, Isaac et al. 1985a). *Figures for the coxII intron are those mtDNA fragments which hybridised only with an intron specific probe from the maize gene (i.e. in addition to those also homologous to the exon probe as shown). The multiple hybridisation signals with the atp9 gene probes are indicated for three enzymes. Multiple hybridisations with this probe were also found with PstI and SalI digested mtDNA (data not shown) nd- not determined.
3.12 Construction of Cosmid Libraries from Fertile and Sterile Sunflower following Partial Digestion of MtDNA.

The usefulness of cosmid vectors for mapping studies relies on the cloning of large sections (up to 40kb) of insert DNA. It was therefore important that conditions for optimising the size of sunflower mtDNA restriction fragments for cloning were defined. Initially, trial digestions of both pBR322 plasmid DNA (Bolivar et al. 1977) and mtDNA with Sau3A were used to assess conditions suitable for subsequent digestion of mtDNA to be used for cloning. For the particular batch of Sau3A enzyme used, 0.025 units per gg of mtDNA for 2 minutes at 37°C was found to give the requisite level of digestion of the sunflower mtDNA. This was visualised as a slight smear below the main mtDNA band upon electrophoresis (Figure 3.1, lane 5). Prolonging the digestion beyond 2 minutes meant that most of the mtDNA fragments were smaller than required (recognised as a more extensive smear, lanes 6 and 7). Cesium chloride gradient fractionation of the partially digested mtDNA allowed selection of fragments in the 30-50kb range for use in the ligation reactions with cosmid vector DNA as described in section 2.2.10.

The cosmid cloning procedure gave a transformation efficiency of 2 x 10^5 cosmids/µg mtDNA from sterile sunflower and of 8 x 10^4 cosmids/µg mtDNA from fertile sunflower. Filter replicates totalling 1000 clones from each library were screened with probes for the \textit{atpA} and \textit{cob} genes. An example of the colony hybridisation results is shown in Figure 3.18. The autoradigraphs shown in (a) and (b) indicate how marked positions on a duplicated filter, probed with \textit{atpA} and \textit{cob} probes respectively, were used to identify those clones which were 'positive' with the different gene probes.

Small scale cosmid DNA preparations of 'candidate' clones identified by the initial colony screen were checked for size by gel electrophoresis. Figure 3.19 shows such a gel where lanes 3-6 show migration rates expected for clones of between 37 and 50kb (i.e. 32kb-45kb mtDNA plus 5kb vector) and comparable to intact λ (lane 1). In contrast, a smaller clone (lane 2) is the probable result of deletion by recombination during the cloning procedure. Cosmid DNA extracted from clones of the correct size were used on DNA dot blots (small amount of sample DNA directly spotted onto a marked nylon filter and then treated as for a Southern blot) for a second screen with the same gene probe. Generally clones were confirmed as positive at the second screen; however 'false positives' (i.e. those identified by the colony hybridisation which did not hybridise on the dot blot) occurred approximately 10% of the time.
Figure 3.18
Cosmid clones of mtDNA from fertile and sterile sunflower indicate that \textit{atpA} and \textit{cob} genes are linked in both genomes
Replicates of cosmid libraries were created by direct transfer of colonies to 3MM filter paper followed by bacterial lysis and fixation of cosmid DNA as described in methods. Filters were then hybridised with probes for \textit{atpA} or \textit{cob} to identify clones containing the region of genomic divergence in sterile sunflower. The filter shown in (a) has been probed with \textit{atpA} and a replicate filter, probed with \textit{cob}, is shown in (b). Two colonies (positions B3 and B5) are positive with both of the probes.

Figure 3.19
Confirmation that Positive Cosmid clones contain a 30-40 kb mtDNA Insert
Cosmid DNA was purified from clones identified by colony hybridisation by Triton lysis as described in methods. Undigested cosmid DNA clones (lanes 2-6) were then checked for an expected size of 35-45kb by electrophoresis in 0.7\% w/v agarose in comparison to intact \( \lambda \) (lane 1). Any clones not of the expected size (e.g. lane 2) were discarded. (Note the low molecular band in lane 4 is RNA which had not been removed from this cosmid DNA preparation).

Table 3.3
Summary of Results of Screening Cosmid libraries for Clones Containing the \textit{atpA} and \textit{cob} genes from fertile and sterile sunflower
The table shows the total number of positive clones, identified by colony hybridisation and confirmed by dot-blot, which contained the \textit{atpA} and \textit{cob} genes from fertile (F) and sterile (S) sunflower. 'Double +' were those clones showing positive hybridisation with both these gene probes.
Table 3.3

<table>
<thead>
<tr>
<th>Gene</th>
<th>atpA</th>
<th>cob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome</td>
<td>F S</td>
<td>F S</td>
</tr>
<tr>
<td>Clones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st screen</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>2nd screen</td>
<td>27</td>
<td>37</td>
</tr>
<tr>
<td>Double +</td>
<td>18</td>
<td>22</td>
</tr>
</tbody>
</table>
Deleted clones resulting from rearrangement during the cosmid cloning procedure occurred only infrequently and all were discarded. The results of colony hybridisation and subsequent confirmation by dot hybridisation starting with 1000 clones from each library are presented in Table 3.3.

After an initial screening of the libraries of mtDNA from fertile and sterile sunflower with the \textit{atpA} and \textit{cob} probes, it was apparent that a number of the same clones gave positive hybridisation with both genes. Comparison of Figure 3.18 (a), screened with \textit{atpA} and (b), screened with \textit{cob}, indicates positive signals at positions B3 and B5 on both filters. This suggested the genes for \textit{atpA} and \textit{cob} are linked at least within 45kb, the cloning capacity of individual cosmids, in mtDNA from both sterile and fertile sunflower.

\textbf{Figure 3.20} shows examples of cosmid clones, positively identified by the \textit{atpA} and/or \textit{cob} probes, which have been fractionated by agarose gel electrophoresis following restriction endonuclease digestion (a). A common vector DNA band of ca.5.4kb is present in clones digested with \textit{EcoRI}, (arrowhead) and can be distinguished from the mtDNA restriction fragments, by hybridisation with labelled pJB8 cosmid DNA if necessary. Representative clones from each library (a clone B17C5 from fertile sunflower and a clone A2H2 from sterile sunflower), were digested with \textit{EcoRI} or \textit{BamHI}, fractionated by electrophoresis, transferred to Hybond-N and hybridised with either a 5' \textit{atpA} probe (ALXR18) (b) or \textit{cob} probe (1435) (c). These hybridisations confirmed that the homologous restriction fragments within these clones were the same as those identified in the genomic mtDNA digests included for comparison.

\textbf{3.13 Detailed Mapping Around \textit{atpA} Indicates a Divergence in the 3' Region of the Gene in Sterile Sunflower}

The 2.2kb and 3.8kb \textit{EcoRI} mtDNA fragments containing the \textit{atpA} gene in the cosmid clones B17C5 and A2H2 respectively were isolated and cloned into the \textit{EcoRI} site of pUC18. Restriction maps of the two \textit{atpA} clones were constructed and are shown in Figure 3.21 with 5' to the left and the extent of homology to the 5' and 3' maize \textit{atpA} probes represented by shading. The clones are referred to as pB22 for the 2.2kb \textit{EcoRI} fragment from fertile sunflower and pA38 for the 3.8kb \textit{EcoRI} fragment from sterile sunflower. The restriction sites at the 5' end of the two clones are identical but of the two \textit{BamHI} sites in the pA38 clone only one is present in the pB22 clone. This second \textit{BamHI} site (boxed in Figure 3.21b) creates an internal 1.2kb \textit{BamHI} fragment in pA38. Part of this 1.2kb \textit{BamHI} fragment is homologous to the 1.5kb \textit{BamHI}-\textit{EcoRI} which constitutes the remainder of the pB22 clone (Figure 3.21a). Most of the additional length of the pA38 clone from sterile sunflower consists of a 1.9kb \textit{BamHI}-\textit{EcoRI} fragment which is not present in the pB22 clone from the fertile line.
Figure 3.20
Restriction Endonuclease Digestion Identifies the atpA and cob Homologous Fragments Within Cosmids Clones
Cosmid clones identified by colony hybridisation with atpA and cob gene probes from each library were used for DNA preparation by Triton lysis as described in methods (a) shows 1μg of five purified cosmid clone DNAs digested with BamHI and XhoI (1), or EcoRI (lanes 2-5), electrophoresed through 0.8%w/v agarose and visualised by ethidium bromide staining. With EcoRI digestion a common 5.4kb pJB8 vector band is apparent (arrowhead). Similar gels were used for transfer of cosmid DNA to Hybond -N for subsequent hybridisation (b) shows an autoradiograph corresponding to hybridisation of representative clones from each library (B17C5 from fertile, A2H2 from sterile) hybridised with the atpA 5’ probe ALXR18 or in (c) to the cob probe 1435. MtDNA from sterile sunflower was included in the hybridisation for comparison. Lanes are (b) 1-mtDNA (sterile S) BamHI, 2-A2H2 EcoRI, 3-A2H2BamHI, 4-B17C5 BamHI 5-B17C5 EcoRI (b) 1-mtDNA (S) EcoRI, 2- A2H2 EcoRI, 3-A2H2 BamHI, 4- B17C5BamHI, 5-B17C5EcoRI.

Figure 3.21 (page 92)
Restriction Endonuclease Mapping of the atpA gene from fertile and sterile Sunflower Indicates a Region of Divergence
EcoRI restriction endonuclease fragments containing the atpA genes from fertile (2.2kb) and sterile (3.8kb) sunflower were subcloned from cosmids B17C5 and A2H2 respectively into the pUC18 plasmid vectors to produce the plasmid clones pB22 and pA38 respectively. These clones were digested with a range of restriction enzymes, the products were separated by electrophoresis on 0.7%w/v agarose gels and visualised by ethidium bromide staining. On the restriction map diagrams produced, shading indicates the minimal extent of homology to atpA in pB22 (a) and pA38 (b) determined by hybridisation with the maize atpA probes ALXR18 (5’) and BLSCI (3’). The letters E, Bg, S, B, H and Sc refer to recognition sites for the enzymes EcoRI, BglII, SalI, BamHI, HindIII and SacI respectively. The SacI digestion does not appear in the photograph but is referred to in the text. All size estimates are in kilobases (kb) and were estimated by comparison with HindIII digested λ DNA (lane 10 and sizes marked at the sides).
Figure 3.20

(a)

(b) atpA

(c) cob
Figure 3.21

(a) fertile pB22

(b) sterile pA38
The maps shown in Figure 3.21 demonstrate that a \textit{Sac}I site, located approximately 1kb from the 5' \textit{EcoRI} site of the 2.2kb and 3.8kb \textit{EcoRI} fragments is common to both clones and indicates that the 5'\textit{atpA} sequence between these two sites is probably identical in fertile and sterile sunflower (precluding the possibility of small mismatches, in sequences other than those of mapped restriction sites). The presence of the diagnostic \textit{Bam}HI site in the pA38 clone from sterile sunflower (boxed in Figure 3.21) indicates that an alteration in the sequence in or near the 3' region of \textit{atpA} is responsible for the generation of this larger fragment which hybridises to the gene probe in this cytoplasm. DNA sequence analysis provided a more detailed analysis of this region of the genome as described in the next chapter.

3.14 Mapping of the \textit{cob} Gene Indicates that it is Closely Linked to \textit{atpA} in the Mitochondrial Genome of Sterile Sunflower

Hybridisation of a \textit{cob} gene probe to total genomic mtDNA indicated that homology is located on different sized \textit{Bam}HI fragments in the fertile and sterile sunflower (Figure 3.6). The 10.1kb and 12.4kb \textit{Bam}HI fragments containing the \textit{cob} gene from fertile and sterile respectively, were subcloned from the cosmids B17C5 and A2H2 into pUC18 for restriction mapping. Figure 3.22 presents the maps of the 10.1kb and 12.4kb \textit{Bam}HI clones (referred to as pB101 and pA124 respectively) with the \textit{cob} coding region shown 5' to the left. An M13 probe specific for the coding region of the maize \textit{cob} gene (M13 640) was hybridised to the clones pB101 and pA124 to locate homology within the 3.6kb \textit{EcoRI} and 3.7kb \textit{HindIII} fragments in both as shown in the autoradiographs in Figure 3.22 (a) and (d) and indicated on the maps by shading. The two \textit{cob} clones show similar restriction maps over most of their length, including the gene coding and the immediate 3' flanking regions. \textit{EcoRI} fragments of 3.6kb and 3.1kb and a 1.4kb\textit{Bam}HI-\textit{EcoRI} fragment at the junction of insert and vector occur in both clones. The pA124 clone from sterile sunflower contains a 2.4kb \textit{EcoRI} fragment and 1.9kb \textit{Bam}HI-\textit{EcoRI} fragment not present in pB101kb. The latter contains a 2.0kb \textit{Bam}HI-\textit{EcoRI} fragment not present on the pA124 clone.

Isolated fragments from the pA124\textit{cob} clone from sterile sunflower, were hybridised individually to \textit{Bam}HI-\textit{EcoRI} double digests of this and the pB101 clone from fertile to confirm that the fragments which appeared to be common to the clones were indeed homologous (Figure 3.22 f-i). This also demonstrated that the 2.4kb \textit{EcoRI} fragment from the clone pA124 hybridised to the 2.0kb \textit{Bam}HI-\textit{EcoRI} fragment of the pB101 clone (f). This 2.0kb \textit{EcoRI}-\textit{Bam}HI junction fragment is part of a 2.1kb \textit{EcoRI} fragment seen in a mtDNA digest of the fertile genome (also shown in f).

The 1.4kb \textit{Bam}HI-\textit{EcoRI} fragment which is common to the two clones is part of a 4.2Kb \textit{EcoRI} fragment present in both genomes (i). No homology to
the 1.9kb BamHI-EcoRI from pA124 is present in the pB101 clone (data not shown). However, this 1.9kb BamHI-EcoRI fragment (marked by an X on the map of the 12.4kb clone) was found to hybridise with the 3.8kb EcoRI atpA clone pA38 from sterile sunflower (shown in e). This 1.9kb BamHI-EcoRI fragment represents the 3' end of the 3.8kb EcoRI fragment beginning at the diagnostic BamHI site described on the map of the pA38 clone (Figure 3.21b). No fragment within the pB101 cob clone hybridised with either the 2.2kb or 3.8kb EcoRI fragments containing the atpA genes.

Thus, the restriction maps of the atpA and cob clones isolated from mtDNA of fertile and sterile sunflower indicate that the linkage of these genes differs. The 12.4kb BamHI mtDNA restriction fragment from sterile sunflower contains both the cob gene and the 3' region of the atpA gene. In contrast, the cob containing 10.1kb BamHI fragment in fertile sunflower contains no atpA homology. The maps for the atpA and cob genes from sterile sunflower indicate that these two coding sequences are orientated in the same direction in this genome. The non-alignment of the plasmid restriction maps of the cob and atpA regions in mtDNA from fertile and sterile sunflower indicates some form of genomic rearrangement in the vicinity of these genes. Cosmid clones covering longer sections of the mtDNA from both genomes, were used to determine the extent of rearrangement and to describe, in more detail, the recombination event(s) responsible.
The 10.1kb and 12.4kb BamHI fragments from cosmids B17C5 and A2H2, which are homologous to the maize cob gene, were isolated and subcloned into pUC18 to produce the clones pB101 and pA124 respectively. These clones were digested with several enzymes, the products were separated by electrophoresis on 0.8%w/v agarose and visualised by ethidium bromide staining before transfer to Hybond-N for subsequent hybridisation. Panel (b) shows the pB101 clone digested with BamHI only (lane 1), then with BamHI and EcoRI (lane 2) BamHI and HindIII (lane 3), BamHI and Sall (lane 4), BamHI and ClaI (lane 5), BamHI and XbaI (lane 6). Panel (a) shows lanes 1-2 from (b) hybridised with an internal cob gene probe from maize (M13 640) which specifically labels the 3.6kb EcoRI fragment within pB101.

Figure (c) shows digestion of pA124 with the same enzymes, in the same order as panel (b). Panel (d) shows lanes 1-2 from (c) probed with cob 640 probe which also hybridises to the 3.6kb EcoRI fragment in this clone. Panel (e) shows the filter from (d) stripped and reprobed with the 3.8kb EcoRI atpA clone (pA38) from sterile sunflower. The 1.9kb BamHI-EcoRI fragment which hybridises is also present in the pA38 clone and is marked by an X on the pA124 map. (The pUC18 vector band in pA124 hybridises strongly, as the entire pA38 clone including the pAT153 vector was labelled for this hybridisation experiment).

In panels f to i, EcoRI digested mtDNA from sterile (lane 1) or fertile (lane 2) sunflower and BamHI-EcoRI digested pA124 (lane 3) and pB101 (lane 4) were fractionated by electrophoresis on 0.8%w/v agarose and transferred to Hybond-N. The individual filters were then hybridised with isolated fragments of the pA124 clone which had been labelled with $^{32}$P. Panel f shows the result with the 2.4kb EcoRI probe. In EcoRI digested mtDNA homology corresponding to the 2.4kb EcoRI fragment in sterile is a 2.1kb EcoRI fragment in fertile. Panel g has been probed with the 3.1kb EcoRI fragment from pA124, h with the 3.6kb EcoRI fragment and i with the 1.4kb EcoRI-BamHI fragment. The latter junction fragment within the two clones correspond to a 4.2kb EcoRI fragment in both fertile and sterile lanes 1 and 2 respectively).
Figure 3.22

Fertile (a, b) B E 5' H E 3' H E C S B
pB101

E 2.0 3.1 3.7 3.6 1.4 1.5
H 4.9 9.1 9.4 1.0
C
S

Sterile (c-e) B E E 5' H E 3' H E C S B
pA124

X

E 1.9 2.4 3.1 3.6 1.4 1.5
H 2.4 3.1 3.7 1.0
C 7.2 11.4 11.7 0.7
S

(a) 

(b) 

(c) 

(d) 

(e) 

(f) 

(g) 

(h) 

(i)

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
3.15 Rearrangement of the Mitochondrial Genome of Sterile Sunflower Involves Both Sequence Inversion and Insertion.

Several cosmids, identified as containing the atpA and cob genes from both fertile and sterile sunflower were used to compare the physical organisation in the region of these genes in both genomes. Cosmid DNA was purified, digested with EcoRI, BamHI or XhoI and the extent of similarity between clones was assessed by comparative gel electrophoresis. Restriction fragments from individual clones were isolated and used as probes to hybridise with other cosmid DNAs to confirm their location within larger fragments and to indicate which fragments were present as junction fragments with the vector in each clone. The restriction maps of the cosmids containing these regions of the genomes in fertile and sterile sunflower are shown aligned in Figure 3.23. Those fragments which are homologous to atpA and cob are shown on the restriction maps of both genomes. A comparison of the two cosmid maps indicates that many of the restriction fragments in the region examined are common to fertile and sterile sunflower. However, the maps do not align in the intergenic region between atpA and cob and in the region immediately 'downstream' of the cob gene.

The data from the restriction mapping of the plasmid clones of the atpA gene (Figure 3.21) indicated that, within sterile sunflower, sequence divergence occurs in the 3' coding or flanking region of this gene. The atpA clone pA38 from sterile sunflower contains a unique 1.2kb BamHI fragment. To locate a fragment within the fertile genome from which the 1.2kb BamHI fragment may have originated, it was isolated for use as a probe against cosmids and total mtDNA from both genomes. Figure 3.24 shows the result of hybridisation of the 1.2kb BamHI fragment to clones representing the two libraries (B17A1 from fertile and A2A5 from sterile) and total mtDNA for comparison. The 2.2kb and 3.8kb EcoRI fragments which contain the atpA genes hybridise with this probe as expected, but additional fragments, of different sizes in the two cosmid clones, homologous to the 1.2kb BamHI fragment are also detected. An 0.8kb EcoRI fragment occurring in clone B17A1 and a 1.2kb EcoRI fragment in clone A2A5 are also present in the mtDNA from the fertile and sterile cytoplasms from which the clones were derived proving that they are authentic genomic fragments.

The mtDNA from both fertile and sterile sunflower show multiple homologies with the 1.2kb BamHI fragment (Figure 3.24). The 0.8kb EcoRI fragment in fertile and the 1.2kb EcoRI fragment in sterile are the only homologous fragments present within the cosmid clones from both genomes which were analysed. This indicates that the other fragments hybridising in the total mtDNA tracks do not occur within the mapped genomic regions. Some of the hybridising fragments (e.g. the 8.5kb and the 2.0kb EcoRI fragments present in both genomes) were identified with the maize 3' atpA probe BLSC1 and may
represent substoichiometric molecules within the genomes (see Figure 3.5). Additional sunflower mtDNA fragments also hybridise with the 1.2kb BamHI probe and appear to be either specific to the sterile genome (e.g. the 7.3kb and 3.0kb EcoRI fragments, indicated by asterisks in Figure 3.24) or common to both genomes (the 0.9kb EcoRI fragment). The hybridisation pattern shown in Figure 3.24 indicates that at least one copy of a repeated sequence, homologous to the 3' region of the *atpA* gene in sunflower, referred to as 'the repeat' for the remainder of this chapter, maps to different fragments within the main genomes of both fertile (0.8kb EcoRI) and sterile (1.2kb EcoRI) sunflower.

The 1.2kb EcoRI fragment which contains the repeat in mtDNA of sterile sunflower was isolated and cloned into pAT153 (this clone is referred to as pA212). This clone was then labelled with $^{32}$P and hybridised to mtDNA in order to confirm that it shows homology with the major *atpA* containing fragments within the two genomes. The pattern of hybridisation with pA212 is shown in Figure 3.25. Comparison with the 1.2kb BamHI hybridisation pattern (Figure 3.24) confirms that a sequence common to the two fragments is repeated several times in mtDNA of both fertile and sterile sunflower.

Cosmid clones containing the repeat from fertile and sterile sunflower were used to map the position of the 0.8kb EcoRI and 1.2kb EcoRI fragments as shown in Figure 3.23. Using the restriction fragments containing the *atpA* gene, the *cob* gene and the repeat as reference points, the mtDNA organisation of this genomic region can be compared between fertile and sterile sunflower. The cosmid maps indicate that the two genomes are identical for over 15kb upstream of the fragments containing the *atpA* gene and for a similar distance downstream of the 0.8kb and 1.2kb EcoRI fragments containing the repeat in fertile and sterile sunflower respectively. Rearrangement between the genomes is thus restricted to the region between the *atpA* genes and the repeats. The fragments which are immediately 3' to those containing the *atpA* genes are not equivalent in fertile and sterile sunflower. A 4.2kb EcoRI fragment is adjacent to the 2.2kb EcoRI *atpA* containing fragment in fertile sunflower. The same 4.2kb EcoRI fragment is not adjacent to the 3.8kb EcoRI *atpA* containing fragment in fertile sunflower, but is found on the 5' side of the repeat in this genome. Similarly, the 2.1kb EcoRI fragment which is 5' to the repeat in the fertile genome, is homologous to the 2.4kb EcoRI fragment adjacent to the 3.8kb EcoRI *atpA* containing fragment in the sterile line. This 2.1kb EcoRI fragment in the fertile genome is part of the mapped 10.1kb BamHI *cob* clone pB101, which is positioned within the genome as shown in Figure 3.23.

98
Figure 3.23
Physical Organisation of the Sunflower Mitochondrial Genome Spanning the Region Encoding \textit{atpA} and \textit{cob}

Cosmid clones from the mtDNA libraries of fertile (i) and sterile (ii) sunflower were digested with EcoRI, BamH or XhoI and fractionated by electrophoresis through 0.8% w/v agarose. Fragments in common between different clones were identified by similar migration and used to assess the extent of 'overlap' between clones. The clones are represented as black bars and their position in the libraries of mtDNA from fertile (suffix B) or sterile (suffix A) sunflower were used as clone names. Restriction fragment sizes are in kilobases. The regions encoding \textit{atpA} and \textit{cob} are indicated (diagonal stripes for \textit{atpA}, shading for \textit{cob} and named). The 0.8kb and 1.2kb EcoRI fragments in the genomes of fertile or sterile sunflower, which contain 'the repeat' are indicated by asterisks on the maps and also by diagonal stripes above the maps. The orientation of the \textit{atpA} homology is indicated by the arrow (points 5' to 3'). The regions of the cosmid maps corresponding to the plasmid \textit{cob} clones pB101 from fertile and pA124 from sterile sunflower are indicated. The 1.2kb BamH1 fragment from sterile sunflower, used as a probe in Figure 3.24 is indicated, as is the 1.2kb EcoRI fragment (clone pA212) used as a probe in Figure 3.25.
Figure 3.23

(i) Fertile

(ii) Sterile

A2H2  A5C3  A12E1  A2A5  A21B5  B19E1  B17C5  B3A1  B17A1  A24E5  pB22  pB101  atpA  cob  repeat  100

BamHI  EcoRI  3'  5'  6.4  5.1  9.7  4.3  10.1  3.5  7.2  3.2  3.65  -1.5  4  9.2  2.2  4.2  3.6  3.1  2.1  8  3.4  3.3  14.9

0.3  3.2  8.8  1.3  5.1  5.7  13.1  16.2

(1.2kb BamHI)  probe  pA212  (1.2kb EcoRI)  repeat

6.4  5.1  9.7  -1.2  12.4  6.6  7.2  3.2  3.65  -1.5  4  9.2  3.8  2.4  3.1  3.6  4.2  1.2  3.4  3.3  14.9

5.7  20.4  8.8  1.3  5.1  5.7  13.1  16.2
**Figure 3.24**

A 1.2kb BamHI Fragment From the *atpA* Gene in Sterile Sunflower Hybridises to Several Other Restriction Fragments

The 1.2kb BamHI fragment from the clone pA38 was labelled with $^{32}$P and hybridised to EcoRI digests of the cosmids A2A5 (lane 1) and B17A1 (lane 2) and mtDNA from sterile (lane 3), fertile (lane 4) and also hybrid (lane 5) sunflower which had been fractionated by electrophoresis through 0.8% w/v agarose gels and then transferred to Hybond-N. The autoradiographs shown are 6 hour exposures for lanes 1-2 and 48 hour exposure for lanes 3-5. In addition to the *atpA* homologous fragments of 3.8kb and 2.2kb, EcoRI fragments of 1.2kb and 0.8kb are labelled in A2A5 and B17A1 respectively. These fragments are among the multiple homologies (indicated by arrows) that this probe identifies in the sunflower mtDNA. Asterisks denote fragments which appear only in the mtDNA from sterile and hybrid lines.

**Figure 3.25**

Hybridisation of pA212 to mtDNA Confirms that this Clone Contains a Sequence repeat of the Genomes of Fertile and Sterile Sunflower

The 1.2kb EcoRI fragment from sterile sunflower cloned in pA212 was labelled with $^{32}$P and hybridised with EcoRI digested mtDNA from fertile (lane 1) and sterile (lane 2) sunflower, fractionated by electrophoresis through 0.8% agarose and then transferred to Hybond-N. The fragments which hybridise are identical to those labelled by the 1.2kb BamHI fragment from pA38 (Figure 3.24) indicating that the 1.2kb EcoRI and 1.2kb BamHI fragments from sterile sunflower contain a sequence which is reiterated elsewhere in the genomes of both the fertile and sterile cytoplasms. The restriction fragments which appear specific to the sterile line (7.3kb and 3.0kb) are indicated by asterisks.
The 12.4kb BamHI cob containing fragment from sterile sunflower, mapped in clone pA124, can also be aligned with the corresponding region on the cosmid map using the atpA homology at the 5' end of this clone to indicate its position and orientation (Figure 3.23). A comparison of this region of the sterile genome with that of the fertile indicates that the order of three consecutive EcoRI fragments (the 4.2kb, the 3.6kb and the 3.1kb) are reversed. In other words, the region between the restriction fragments which contain the atpA gene and the repeat in sterile sunflower, a distance of approximately 13kb, is inverted with respect to the corresponding region of the genome in fertile. As the endpoints of this inversion are located in fragments which are at least partially homologous, (i.e. the 3.8kb and 1.2kb EcoRI fragments) this suggests that this rearrangement in sterile sunflower could have originated by homologous recombination.

In the genome of fertile sunflower, recombination between a sequence present in the 'parental' 2.2kb EcoRI fragment and indirectly repeated in the 0.8kb EcoRI fragment would predict an inversion of the section between them as is found to occur in sterile sunflower. The fragments which flank those containing the repeat in both genomes are single copy sequences (shown earlier for the 2.1kb EcoRI fragment in the fertile genome, Figure 3.22f, data not shown for the 3.4kb EcoRI fragment). This indicates that the repeat in the 0.8kb and 1.2kb EcoRI fragments is short and does not extend beyond these fragments in the respective mitochondrial genomes from which they are derived.

The cosmid and plasmid mapping data do not accurately locate the endpoints of the inversion within the genome of sterile sunflower, however they do demonstrate the consequences for the organisation of the gene sequences involved, as depicted in Figure 3.26. To produce the genomic inversion through homologous recombination, the repeat within the fertile genome must be indirect with respect to the main atpA homology. The DNA hybridisation analyses do not define the exact nature of the sequence within the repeat. If this repeat contains homology to the coding region of atpA, it is possible that the recombination which has occurred in sterile sunflower may have included part of the atpA gene in this genome. Alternatively, recombination may have occurred downstream of the coding sequence to produce divergence in the 3' flanking region. As the genomic inversion in sterile sunflower completely spans the region encoding the cob gene, this sequence is inverted with respect to the atpA gene in the mtDNA from this cytoplasm (Figure 3.26).

When the distance between the restriction fragments which contain the proposed inversion endpoints in sterile sunflower (3.8kb and 1.2kb EcoRI fragments) and corresponding regions in fertile sunflower (2.2kb and 0.8kb EcoRI fragments) are compared, there is a difference in size of 2.3kb. This indicates that at least one insertion event has occurred in the inverted region.
In the genome from sterile sunflower, the restriction pattern across the 3.8kb and 2.4kb EcoRI fragments does not align with the corresponding 4.2kb and 0.8kb EcoRI fragments in fertile sunflower, as expected for a simple inversion. The non-alignment at this 5' end of the inversion suggests this as the likely position of the 2.3kb insertion.

This region presumed to have arisen by insertion therefore includes the 3' end of the 3.8kb EcoRI atpA containing fragment from sterile sunflower cloned in pA38. In order to investigate the origin of the sequence insertion in sterile sunflower, part of this clone was used for further hybridisation experiments. The 1.9kb BamHI-EcoRI fragment from the 3' end of pA38 was isolated and hybridised to mtDNA from both fertile and sterile sunflower. The 1.9kb probe labelled a single BamHI fragment of 11.2kb in mtDNA from fertile sunflower and a 10kb BamHI in addition to the 12.4kb BamHI fragment it is part of in mtDNA from sterile sunflower (Figure 3.27). Thus sequences homologous to the region of inserted DNA exist elsewhere in the fertile genome and this suggests an intragenomic recombination event may be the origin of the insertion in the genome of sterile sunflower. No hybridisation with the 1.9kb BamHI-EcoRI probe was detectable in a chloroplast DNA preparation from fertile (lane 1) or sterile sunflower (lane 2 in Figure 3.27) as further support that the inserted DNA is derived from elsewhere in the mitochondrial genome.

It was of note that one cosmid, isolated from the library of mtDNA from sterile sunflower, was found to contain an organisation identical to that of the fertile line. The cosmid A24E5 contained the 2.2kb and 0.8kb EcoRI atpA homologous fragments and a total of 37.5kb of mtDNA identical to that of clones from the library of fertile sunflower mtDNA. This cosmid is presumably representative of the subliminal level of the fertile type organisation which was identified in the sterile line during the hybridisation analysis.

The physical maps of the fertile (HA89B) and sterile (CMS89) sunflower mitochondrial genomes described by Siculella and Palmer (1988) were virtually identical to each other. The maps are best aligned by postulating an inversion of ca. 12kb downstream of the atpA gene. The genome of the sterile CMS89 line was reported to contain an insertion of 5kb at one end of the inversion. This is larger than the insertion in the mtDNA of the sterile line 842A examined here and suggests that recombinational insertion into the mtDNA of these two sterile cytoplasms may have involved different sequences. Unfortunately, no description of the repeats involved in genomic inversion in the CM89 line has been reported thus precluding a more detailed comparison with the 842A rearrangement described in this thesis.
Figure 3.26
Rearrangement of the Mitochondrial Genome from Sterile Sunflower involves both Sequence Inversion and Insertion
The data from plasmid and cosmid mapping studies is summarised as a diagrammatic representation of the mitochondrial genome rearrangement in sterile sunflower when compared to fertile. The \textit{atpA} coding and 3' region is represented by diagonal lines and is shown with the coding strand (upper line) 5' to the left. The \textit{cob} gene, on the complementary strand to \textit{atpA} in fertile sunflower, is on the same strand in the sterile sunflower. The proposed indirect repeat of the \textit{atpA} 3' coding and/or flanking region in the fertile genome is also represented by diagonal lines. Recombination between these two repeated sequences would result in the genomic inversion observed in the sterile sunflower genome, shown diagrammatically at the bottom of the figure. In addition, a sequence insertion of approximately 2.3kb adjacent to the 3' end of the \textit{atpA} gene in sterile sunflower has also occurred (shaded black) The restriction fragment sizes shown are those produced following digestion with \textit{EcoRI} (see Figure 3.23). Vertical lines join homologous restriction sites in the two genomes.

---

\begin{itemize}
\item \textit{atpA} \hspace{1cm} \textit{repeat} \hspace{1cm} \textit{insertion} \hspace{1cm} \textit{cob} (5' to 3')
\end{itemize}

Figure 3.27 (page 109)
The Region of Sequence Insertion Near \textit{atpA} in the Sterile Sunflower mtDNA has Probably Arisen by an Intragenomic Recombination Event

Chloroplast DNA from sterile (lane 1) and fertile (lane 2) sunflower and mtDNA from sterile (lanes 3 and 4) and fertile (lane 5) sunflower was digested with \textit{BamHI} (lanes 1-2, 4-5) or \textit{EcoRI} and \textit{BamHI} (lane 3), electrophoresed through 0.8% w/v agarose and transferred to Hybond-N. This filter was hybridised with a $^{32}$P labelled probe of the 1.9kb \textit{EcoRI-BamHI} fragment from clone pA38, containing the region of possible sequence insertion in sterile sunflower.
3.16 A Novel Mitochondrial Plasmid is Amplified in Restorer Sunflower

The presence of the 1.45kb P₁ plasmid-like molecule in mtDNA of fertile sunflower was described earlier. Neither this plasmid molecule nor the related 1.8kb P₂ found in some other cytoplasms, hybridises with the main band mtDNA (Perez et al. 1988). During the course of mtDNA hybridisation analysis with the pA212 clone, which contains the 1.2kb EcoRI 'repeat' fragment from the main genome of sterile sunflower (see Figure 3.23), it became apparent that another mitochondrial 'plasmid-like DNA' was present in the restorer (843R) line of sunflower. Upon gel electrophoresis and visualisation by ethidium bromide staining, the 'plasmid' migrates at a rate corresponding to a molecular weight of 5kb in a preparation of native mtDNA from the restorer line (Figure 3.28). However, if the 'plasmid' is of a closed circular form this is an underestimate of the size due to the unusually high rate of migration of covalently closed circles in agarose gels (Old and Primrose 1985). A stained band is also apparent which migrates more slowly than the main mtDNA and may correspond to a multimeric form of the monomeric 5kb molecule. Alternatively, it may represent the 'plasmid' specifically bound to copurifying proteins although the proteinase treatment of the mitochondria used to release the mtDNA means it is unlikely that any such proteins would still be present and lead to a retarded migration in agarose.

Figure 3.28 indicates the level of the '5kb plasmid', relative to the main mitochondrial genome in the restorer line which was estimated to be present at 1-2 copies per genome by comparison to the plasmid P₁ (see Figure 3.1). However, the level of hybridisation to the '5kb plasmid' noted in the EcoRI and BamHI digested mtDNAs in the mtDNA from restorer sunflower is much greater than a two-fold increase relative to genomic fragments as expected for this copy number (Figure 3.29). An explanation for this may be that the plasmid DNA contains a large number of repeats, of the part of the 1.2kb EcoRI sequence to which it is homologous. BamHI digestion of mtDNA from the restorer line does not alter the pattern of migration of the '5kb plasmid' from that seen with native mtDNA, suggesting that it does not contain a recognition site for this enzyme. In EcoRI digested mtDNA, hybridisation due to the 'plasmid' is located mainly on a 3.7kb fragment. The 6kb EcoRI fragment which hybridises faintly may represent some partially digested multimeric form of the 'plasmid'. The fact that the 'plasmid' shares homology with the main band mtDNA and appears to contain repeated copies of this homology suggests it may represent a sub-genomic molecule which has become specifically amplified in the cytoplasm containing the restorer nucleus. No further characterisation of this molecule was undertaken but further proposed investigations are discussed in Chapter 7.
Figure 3.28
The Mitochondrial Genome of Restorer Sunflower Contains a 'plasmid-like' DNA Molecule

4μg of native mtDNA from the 843R restorer line of sunflower (lane 1) was electrophoresed on an 0.8% w/v agarose gel and visualised by ethidium bromide staining to demonstrate the presence of a low molecular weight band, separate from the main genomic mtDNA. By comparison with HindIII digested λ DNA (lane 2) the estimated molecular weight of the undigested molecule is 5kb (this may be an underestimate if the molecule is a closed circular plasmid-like DNA similar to P1).

Figure 3.29
The 'plasmid-like DNA' is Homologous to the Clone pA212 Containing the atpA 3' Repeat from Sterile Sunflower

3μg of mtDNA from the sterile line 841A (lanes 1 and 7), the fertile line 841B (lanes 2 and 8), the sterile line 842A (lanes 3 and 9), the fertile line 842B (lanes 4 and 10), the hybrid line 844H (lanes 5 and 11), and the restorer line 843R (lane 6 and 12) were digested with EcoRI (1-6) or BamHI (7-12) and the products separated by electrophoresis through 0.8% w/v agarose before transfer to Hybond-N. This filter was hybridised with the clone pA212 from sterile sunflower which had been labelled with 32P. The figure shows an autoradiograph of the blot after 4 hours exposure and indicates the greatly amplified signal in restorer sunflower line in comparison with the other lines. The reduced size of the hybridising fragment following EcoRI digestion indicates that at least one recognition site for this enzyme occurs within the 'plasmid' sequence. BamHI digestion mtDNA shows a major hybridising signal corresponding to the size of the 'plasmid' seen in gel electrophoresis of native mtDNA (i.e. 5kb) indicating that it contains no site for this enzyme. The larger hybridising fragments in the BamHI samples therefore most likely represent the multimeric forms of this molecule.

Relative to the restorer line a very low level of this 'plasmid' was found to be present in cytoplasts from the fertile (842B), sterile (842A) and hybrid sunflower (844H) when the 1.2kb EcoRI fragment was used to probe undigested mtDNA from these lines (data not shown)
3.17 Conclusions Regarding Mitochondrial Genome Organisation in Fertile and Sterile Sunflower

Analysis of the organisation of the mitochondrial genomes of fertile and sterile sunflower by hybridisation with specific gene probes demonstrated a number of features of interest with regard to coding capacity, gene duplications, linkages and gene structure. These are discussed below in relation to mitochondrial genome organisation in several other species. Hybridisation analysis also provided evidence for reorganisation between the mitochondrial genomes of fertile and sterile sunflower associated with the genes encoding the α-subunit of the F$_1$ATPase (atpA) and the apocytochrome subunit b (cob).

Sunflower mtDNA was found to contain sequences homologous to the atpA gene, represented at substoichiometric levels and some of which are specific to the genome from the sterile cytoplasm. An indication that amplification of particular subgenomic molecules may occur within the different cytoplasms of sunflower was provided by the finding of a 5kb 'plasmid-like' DNA molecule which is highly amplified in mtDNA from sunflower containing a restorer nucleus.

Restriction endonuclease mapping of plasmid and cosmide clones of mtDNA from fertile and sterile sunflower have shown that a major rearrangement involving inversion of a ca. 13kb region of the mitochondrial genome of sterile sunflower has probably occurred as a result of homologous recombination between the 3' end of the atpA gene and a short, indirect repeat. The region of the genome which has been inverted in sterile sunflower spans the cob gene and may also include part of the atpA coding region. In addition, the 5' end of the inversion contains approximately 2.3kb of sequence not present in the corresponding position of the mitochondrial genome of fertile sunflower.

A number of testable predictions arise from the observations presented so far concerning the atpA-linked rearrangement between fertile and sterile mtDNA. These are as follows:

a) the 2.2kb and 3.8kb EcoRI fragments which contain the atpA genes in fertile and sterile sunflower respectively diverge in sequence only after a common SacI site located approximately 1000bp from their 5' EcoRI sites;

b) the EcoRI fragments of 0.8kb (in fertile) and 1.2kb (in sterile) will contain sequence homology to the 2.2kb and 3.8kb EcoRI fragments in the form of indirect repeats of the 3' region of the atpA gene;

c) due to the insertion in the mitochondrial genome from sterile sunflower a second sequence divergence will occur in the 3' region of the atpA gene in this cytoplasm.

Sequence analysis of the atpA genes and associated repeated sequences allowed a more accurate definition of the probable sites of recombination responsible for the rearrangements in sterile sunflower (see Chapter Four).
In addition to the predictions, a number of questions were posed by the findings described. The most important is whether the rearrangement found in sterile sunflower is of consequence with respect to the expression of either the \textit{atpA} or \textit{cob} genes in the mitochondria from this cytoplasm. As already stated, only variation which is expressed can be responsible for phenotypic consequences such as male-sterility. Many examples of mitochondrial genome reorganisation, due to aberrant recombination, have been reported in other plant species without any link being established to phenotypic expression (discussed below). Therefore, it was necessary to investigate whether the differences in mtDNA organisation found between fertile and sterile sunflower were expressed as changes in transcription patterns between the cytoplasms.

\textbf{3.18 Mitochondrial Genome Heterogeneity in Higher Plants}

Restriction endonuclese digestion of sunflower mtDNA reveals a complex profile consisting of a large number of fragments, at variable stoichiometries. This pattern is indicative of a heterogeneous genomic structure probably consisting of subgenomic circles but which can also be represented as a complete 'master' circle. Such subgenomic molecules, described in a number of higher plants, are probably generated by homologous recombination between short repeated sequences and maintained by replication which is coordinated with that of the main genome (Leaver \textit{et al.} 1988). Thus, the genomic organisation of higher plant mtDNA \textit{in vivo} is not fixed and will be determined by the number of actively recombining repeated sequences which are present and the efficiency of different replication origins contained within the different subgenomic circular molecules. To what extent the actual composition of the sequence of repeats influences the frequency of recombination in plant mtDNA is still uncertain.

It is now apparent that, in addition to the relatively frequent recombination events leading to readily identifiable recombination products within plant mitochondrial genomes, there are instances where recombination occurs less frequently. This can lead to very low levels of less easily detectable genomic organisations as described for maize by Small \textit{et al.} (1987) or may be responsible for novel genomic rearrangement events in which reciprocal recombination products cannot be identified. In some cases such novel recombination processes have resulted in the creation of chimeric open reading frames, which are expressed as variant polypeptides and some of which are implicated in the CMS phenotype (section 1.9). In sunflower there are indications that such infrequent recombination processes have contributed to the observed differences in the organisation of the mitochondrial genome and are responsible for variation between the fertile and sterile cytoplasmic types.
As described in section 3.5 of this chapter, a DNA probe of the 3' region of the \textit{atpA} gene hybridised to a number of fragments in \textit{EcoRI} digested mtDNA from fertile and sterile sunflower. Multiple homologies to the \textit{atpA} gene have also been observed within the mitochondrial genomes of maize and \textit{Oenothera} (Small \textit{et al.} 1987, Schuster and Brennicke 1986). In mtDNA from maize alternative forms of the \textit{atpA} gene have been detected at variable levels within and between different cytoplasmic types. The term 'sublimons' has been used to describe the substoichiometric molecules present within the genomes which are responsible for the hybridisation patterns observed (Small \textit{et al.} 1987).

In \textit{Oenothera} mtDNA, in addition to a single intact reading frame encoding \textit{atpA}, at least three pseudogenes are also present (Schuster and Brennicke 1986). At least one of the \textit{atpA} pseudogenes occurs in several locations in \textit{Oenothera} mtDNA within different subgenomic circular molecules. The repeated \textit{atpA} sequences are flanked by unique sequences on either side. Homologous recombination across the 'repeated core' in the intact gene and one of the pseudogenes can explain the origin of two other \textit{atpA} 'alleles' although the precise location of recombination boundaries is complicated by multiple rearrangements. Apparently, only the complete copy of \textit{atpA} is transcribed in \textit{Oenothera}.

Thus, in at least two other species sequences in and around the gene encoding \textit{atpA} are recombinogenic and lead to alternative organisations of those regions of the genome containing this gene. Whether the restriction fragments which hybridise, at low level, to the \textit{atpA} gene in sunflower mtDNA are due to homology to the flanking region or to pseudogenic \textit{atpA} sequences remains to be shown.

Another indication of sequence repetition in the sunflower mitochondrial genome was found during hybridisation analysis with a probe containing the \textit{atp9} gene (section 3.6). MtDNA from fertile and sterile sunflower show similar \textit{atp9} hybridisation patterns, and indicate that at least two different genomic locations are homologous to this gene. Two complete copies of \textit{atp9} have been found in the mitochondrial genome from a fertile \textit{Petunia} cytoplasm. These genes diverge immediately 5' and 3' to the \textit{atp9} coding sequence and thus are not part of some longer repeat within the genome but rather represent a specific gene duplication (Rothenberg and Hanson 1987). A third \textit{atp9} homology is also present in this cytoplasm which has not yet been characterised (Hanson \textit{et al.} 1988). In sterile \textit{Petunia} in addition to a complete copy of the \textit{atp9} gene, the chimeric \textit{Pcf} gene, implicated in CMS (section 1.9.5) is derived, in part, from the 5' region of \textit{atp9} (Young and Hanson 1987).

Two complete copies of the \textit{atp9} gene are also present in mtDNA of the fertile and \textit{Ogura} (sterile) cytoplasms of \textit{Brassica napus} (Markaroff and Palmer 1988). Inversions within the mitochondrial genome of the \textit{Ogura} cytoplasm have translocated the \textit{atp9} genes to different genomic locations in this mtDNA.
The rearrangements do not appear to have consequences for expression of the gene in the two different cytoplasms and a similar, single homologous \textit{atp9} transcript is found in both. Nevertheless, these mtDNA observations suggest that sequences within or around the \textit{atp9} coding region are recombinogenic, leading to duplication and insertion of the gene at multiple locations within mitochondrial genomes of a number of species.

The gene for \textit{atp6} was found to be single copy and identical in mtDNA from sterile and fertile sunflower (section 3.6). In a number of other plant mtDNAs \textit{atp6} homologous sequences have been found in conjunction with genomic rearrangements suggesting that, like \textit{atp9}, this gene appears to be associated with recombinogenic sequences in plant mtDNA. For example in the soybean mitochondrial genome there are two copies of the \textit{atp6} gene which are identical in their central and 3’ regions but differ in the organisation at their 5’ ends (Grabau \textit{et al.} 1988). One \textit{atp6} gene contains homology to the 5’ end of the soybean \textit{coxII} gene including the sequence for the 18 N-terminal amino acid residues of the \textit{coxII} protein. The \textit{coxII} sequence is fused in frame with the \textit{atp6} coding region and contains an ATG codon which, if used for initiation, would result in the synthesis of a chimeric protein. The second \textit{atp6} gene in soybean is fused at its 5’ end to an \textit{atp9} homologous sequence which includes the initiation codon for \textit{atp9} translation. A chimeric transcript corresponding to the \textit{atp6-coxII} gene fusion is the only abundantly transcribed form of the gene in soybean (Grabau \textit{et al.} 1988).

In CMS-T maize, a region homologous to the 5’ flanking sequence of \textit{atp6} which contains the promoter is responsible for expression of the \textit{Turf13} gene implicated in male sterility (section 1.9.4). Also, in CMS-C maize, the 5’ region of the \textit{atp6} gene contains several rearrangements compared to the gene from the fertile (N) cytoplasm (Dewey \textit{et al.} 1985). In sunflower mtDNA from the fertile and sterile cytoplasms, there are no indications of genomic rearrangement involving \textit{atp6} homologous sequences providing further support that the organisation of the mitochondrial genomes is very similar.

The finding of an intron in \textit{coxII} in sunflower mtDNA (section 3.7) was unusual for a dicotyledonous plant but was subsequently confirmed by Gallerani \textit{et al.} (1988) who sequenced the entire gene. The first \textit{coxII} genes to be sequenced from dicotyledonous species were pea and \textit{Oenothera} neither of which were found to contain an intron (Hiesel and Brennicke 1983, Moon \textit{et al.} 1985). This contrasted with the situation in the monocotyledons maize, rice and wheat, all of which have 'interrupted' \textit{coxII} genes (Kao \textit{et al.} 1984, Bonen \textit{et al.} 1988). However, the findings of \textit{coxII} intron sequences in the mitochondrial genomes of sunflower and another dicot, \textit{Petunia}, recently reported by Hanson \textit{et al.} (1988) indicates that this correlation no longer holds.
In sunflower mtDNA only single copy homology to a probe from the 3' region of the \textit{nad1} gene from \textit{Oenothera} was detected (section 3.8). Homology to the 5' region of this gene, on the basis of comparison with fungal mtDNA sequences, has not been located in any plant mitochondrial genome. In tobacco mtDNA a single region of \textit{nad1} homology is present which contains the 3' region of the gene but a truncated first exon encoding only 13 amino acid residues homologous to \textit{Aspergillus nad1} gene, none of which encode methionine to initiate translation (Bland \textit{et al.} 1986). Similar truncated versions of \textit{nad1} are present in maize, watermelon and soybean (Bland \textit{et al.} 1986, Stern \textit{et al.} 1986, Wahleithner and Wolstenholme 1988b). For the moment therefore, the function of \textit{nad1} homologous sequences in plant mtDNA remains unclear. They are not present in all plant species and are apparently present but untranscribed in some others. An investigation of \textit{nad1} transcription in sunflower mitochondria is described in Chapter 5.

In conclusion, a striking similarity of the pattern of fragments produced by restriction endonuclease digestion of mtDNA from fertile and sterile sunflower cytoplasms, coupled with the similarity of the restriction endonuclease fragments identified with gene specific probes suggests that the genomic rearrangement associated with the \textit{atpA} gene stands a high possibility of being causally associated with the CMS phenotype.
4.1 Aims and Rationale

In order to define the point of divergence within or near the 3' end of the \textit{atpA} coding region from fertile and sterile sunflower, a comparison of the nucleotide sequence of these genes was required. Sequence analysis of regions showing homology to the 3' region of the \textit{atpA} gene was also necessary to identify the extent of sequence reiteration in the two genomes and to provide a detailed description of the putative sites of recombination in the \textit{atpA}-linked rearrangement in sterile sunflower. Sequence analysis of the 5' and 3' flanking regions of the \textit{atpA} genes from fertile and sterile sunflower could also increase our knowledge of those sequences involved in the control of expression of plant mitochondrial genes. Thus, sequencing of the \textit{atpA} gene and associated rearrangements in the mitochondrial genome of sunflower was of interest, specifically with respect to the CMS phenotype in this species, and also to increase our knowledge of mitochondrial gene structure.

4.2 Strategy for Sequencing the \textit{atpA} gene from Fertile and Sterile Sunflower

The 2.2kb and 3.8kb \textit{EcoRI} fragments, which contain the \textit{atpA} gene from fertile and sterile sunflower respectively, were used to generate short overlapping single-stranded M13 clones for sequence analysis. The sequencing strategy which was used is outlined in Figure 4.1. Restriction fragments of 200-400bp produced by \textit{TaqI} and \textit{Sau3A} digestion of the 2.2kb and the 3.8kb \textit{EcoRI} fragments were cloned into the vector M13. The resulting libraries were screened for clones corresponding to specific regions of the original 2.2kb \textit{EcoRI} fragment (i.e. the 0.7kb \textit{EcoRI-BamHI} and the 1.5kb \textit{BamHI-EcoRI} fragments) and the 3.8kb \textit{EcoRI} fragment (i.e. the 0.7kb \textit{EcoRI-BamHI}, the 1.2kb \textit{BamHI} and the 1.9kb \textit{BamHI-EcoRI} fragments, see Figure 3.21). For analysis of the 5' flanking sequence of the \textit{atpA} gene, \textit{HindIII} generated mtDNA fragments of sterile (9.3kb) and fertile (5.3kb) sunflower were used to isolate \textit{Sau3A} clones spanning the upstream region in both genomes. A 'universal' M13 oligonucleotide primer was used to prime sequencing reactions with the short \textit{TaqI} and \textit{Sau3A} clones. Chemically synthesised oligonucleotides designed to recognise and prime at specific sequences within a 2.7kb \textit{SacI-EcoRI} subclone of the 3.8kb \textit{EcoRI} fragment and a 1.1kb \textit{SacI-EcoRI} subclone from the 2.2kb \textit{EcoRI} fragment were used to confirm the sequences within the region of divergence between the two genomes (the position of these oligonucleotides is marked by asterisks in Figure 4.1).
Figure 4.1
Strategy for sequence analysis of the atpA gene from fertile and sterile sunflower.
The 2.2kb and 3.8kb EcoRI fragments containing the atpA genes from fertile and sterile sunflower (from clones pB22 and pA38 respectively) were digested with Sau3A (S) or TaqI (T) to generate 200-400bp fragments. These shorter fragments were cloned into the BamHI site (for Sau3A fragments) or the AccI site (for TaqI) of mp18 and mp19 M13 vectors. The HindIII fragments containing atpA in fertile (5.3kb) and sterile (9.3kb) were used to generate the Sau3a clones covering the 5' region of the gene which includes the 5' EcoRI site. Sequenced restriction sites, noted in the mapping analysis are indicated (S-SalI, B-BbsI, C-ClaI, B-BamHI and Sc-SacI). The direction of the arrows indicates the orientation of individual single-stranded clones within the larger fragments. The asterisks mark the position of specific oligonucleotides (all 17 nucleotides in length) which were used to prime sequencing reactions with clones covering the region of divergence between fertile and sterile sunflower. An indication of the atpA coding (black) and non-coding (white) regions which were sequenced are indicated in the lower part of the diagram. The region of sequence divergence within the pA38 clone of the 3.8kb EcoRI fragment from mtDNA of sterile sunflower is indicated by stippled shading.
The sequence of the *atpA* coding region and the immediate 5' and 3' flanking regions from fertile sunflower is presented in Figure 4.2a. The predicted protein sequence, derived from the *atpA* open reading frame (ORF) in fertile sunflower, is also shown. The predicted ORF is identical to the *atpA* gene sequence determined from sterile sunflower. In Figure 4.2a, the *atpA* coding sequence from sunflower (S) has been aligned, for comparison, with the *atpA* genes from maize (M) and *Oenothera* (Oe) (Braun and Levings 1985, Schuster and Brennicke 1986).

4.3 Nucleotide Sequence Analysis of the *atpA* Gene from Fertile and Sterile Sunflower

The *atpA* gene from sunflower (both fertile and sterile) is 92% and 93% homologous at the nucleotide level, to the genes in maize and *Oenothera* respectively. These values are typical of the level of sequence conservation reported for other protein coding genes in higher plant mitochondria. Homology between the sunflower sequence and maize and *Oenothera atpA* sequences begins at the ATG predicted to be the start of translation in these genes. An internal methionine, encoded by bases 52-54 into the *atpA* open reading frame in maize (marked by the asterisk in Figure 4.2a) is preceded by a sequence 5' GGAACTCA 3' (29 nucleotides upstream of this ATG in the maize sequence) which shows some homology with the putative ribosome binding sites of other mitochondrial genes (Dawson et al. 1984). In contrast, the first ATG in the maize ORF is not preceded by any sequence related to the postulated binding site. Despite this, it has been designated the probable *in vivo* initiation codon on the basis of comparison of the predicted amino acid sequence of the F\(_1\) α-subunit polypeptides from other organisms (see Figure 4.4, page 128), and because the gene sequence from *Oenothera* contains no equivalent internal ATG codon. In the sunflower *atpA* gene, as in *Oenothera*, the 52-54 position is occupied by an AU (isoleucine) codon increasing the likelihood that the designated maize ATG is indeed the functional start of translation of all three genes.

The nucleotide sequences upstream of the *atpA* ORFs from the three plant species diverge almost immediately and make meaningful alignments difficult. However, at -42 to -32 relative to the ATG in sunflower there is a pyrimidine-rich cluster 5' TTTGTTCTTTG 3' which is an 8 out of 11 match to sequences located at -56 to -46 and -67 to -57 upstream of the ATGs of the maize and *Oenothera* genes respectively (underlined in Figure 4.2a). A similarity between these sequences and those reported for mapped 5' termini of the shortest transcripts corresponding to the maize *cox1* and the *Petunia atp9-1* gene is indicated by the alignment shown in Figure 4.3 (page 124).
Figure 4.2
Nucleotide Sequence and Predicted Amino Acid Sequence of the atpA gene from Fertile and Sterile Sunflower

(a) The open reading frame corresponding to the atpA gene in sunflower (S) (the sequences from fertile and sterile sunflower are equivalent in this region so no distinction is made) is aligned with the genes from maize (M) and Oenothera (Oe) with the putative methionine initiation codon (ATG) and stop codons (TAA, TGA and TAG in sunflower, maize and Oenothera respectively) shown in bold type-face (Braun and Levings 1985, Schuster and Brennicke 1986). The coding strand of the mtDNA is shown and the predicted amino acid sequence of the sunflower gene is included immediately below the nucleotide sequence from this species. Nucleotides in the sunflower sequence are numbered with respect to the A (+1) of the initiator ATG codon. Within the 5' flanking region, the consensus transcription initiation or processing site, discussed in the text, is underlined and the putative ribosome binding site is indicated by the dotted overline.

Panels (b) and (c) (page 123) show sequencing gels used for the electrophoretic separation of products of dideoxy-chain terminating reactions used for sequence analysis of the atpA genes from fertile and sterile sunflower.

(b) Shows an autoradiograph of a 6%v/v polyacrylamide gel used to electrophorese the products of a sequencing reaction with a TaqI clone containing the coding region of the atpA gene from sterile sunflower beginning at nucleotide position +908 (reading from bottom to top gives the sequence in the 5'-3'orientation as indicated).

(c) Shows an autoradiograph of a 6%v/v polyacrylamide gel after electrophoretic separation of sequencing products of two M13 clones covering the point of sequence divergence in the atpA 3' region, between fertile and sterile sunflower. This shows the point mismatch (marked by asterisks) between the 3' region of the atpA gene of sterile sunflower (a G at position +1691 reading from bottom to top is indicated within the sequence beginning with six G residues at +1671) and of fertile sunflower (an A at position +1691, reading from top to bottom). Complete divergence of the two sequences occurs from position +1728 and is indicated by the arrowheads.
<table>
<thead>
<tr>
<th>Start Line</th>
<th>End Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>586</td>
<td>645</td>
<td>ThrSerGluSerGluThrTyrCysValTyrLeuAlaAlaGlyGlnlysArgSerThr</td>
</tr>
<tr>
<td>646</td>
<td>705</td>
<td>ValAlaGlnLeuValGlnIleLeuSerGluAlaAsnAlaMetGluTyrSerIleLeuVal</td>
</tr>
<tr>
<td>706</td>
<td>765</td>
<td>AlaAlaThrThrAspProAlaProLeuGlnPheLeuAlaAlaProTyrGlyCysAla</td>
</tr>
<tr>
<td>766</td>
<td>826</td>
<td>MetGlyGluTyrPheArgAspAsnGlyMetHisAlaLeuIleTyrAspAspLeuSer</td>
</tr>
<tr>
<td>826</td>
<td>886</td>
<td>LysGlnAlaValAlaTyrArgGlnMetSerLeuLeuLeuArgArgProGlyArgAsp</td>
</tr>
<tr>
<td>886</td>
<td>946</td>
<td>GlyPheProGlyAspValPheTyrLeuHisSerArgLeuLeuGluArgAlaAlaLysArg</td>
</tr>
<tr>
<td>946</td>
<td>1006</td>
<td>SerAspGlnThrGlyAlaGlySerThrAlaAlaProValIleGluThrGlnAlaGly</td>
</tr>
<tr>
<td>1006</td>
<td>1066</td>
<td>TCGGACCAGACAGCGCCAGGGATGCTGACCCCTCTACCCCTCATGGAAACACAGCTGGA</td>
</tr>
<tr>
<td>1066</td>
<td>1126</td>
<td>AspValSerAlaTyrIleProThrAsnValIleProThrAspGlyGlnIleCysSer</td>
</tr>
<tr>
<td>1126</td>
<td>1186</td>
<td>GluThrGluLeuPheTyrArgGlyIleArgProAlaIleAsnValGlyLeuSerValSer</td>
</tr>
<tr>
<td>1186</td>
<td>121</td>
<td>ArgAlaGlySerAlaAlaGlnLeuLysThrMetLysGlnValCysGlySerSerLysLeu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GluLeuAlaGlnTyrArgGluValAlaAlaAlaGlnPheGlySerAspAlaAAspAla</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAAATTGCGCACAATATCGCCAGATGCGCCGAGGTCCCTCTCGTATTGGTGAGCTGACCTGATGCT</td>
</tr>
</tbody>
</table>
1246  GCCGACTCAGGCATTACTCAATAGAGGTGCAAGGCTTACAGAAGTGCCCAGACCAACCACAA
AlaThrGlnAlaLeuLeuAsnArgGlyAlaArgLeuThrGluValProArgGlnProGln
1306  TATACGCCACTTCAATTGAAACAAATTGGATATCTATTTATGCAAGCTGTAATGGATTTC
TyrThrProLeuProIleLeuLysGlnIleLeuValIleTyrAlaAlaValAsnGlyPhe
1366  TGTGATCGAATGCCACTAGACAGAATTTCTCAATATGAGAGAGCCATTCCACAAAGTGTA
CysAspArgMetProLeuAspArgIleSerGlnTyrGluArgAlaIleLeuLysSerIle
1426  AAAACAGAAATTACTAATACCCCTTTTAAAGATGGTGCTTTAAGCTAACGAAAGAAATG
LysThrGl uLeuLeuGlnSerLeuLeuGluGlyLeuThrAsnGluArgGlyMet
1486  GACACGACTACATTCTAAGGAATGCGGCTTTGCTTACACAAATATAAAAAGAAAAAGA
GluProAspThrPheLeuLysGluCysAlaLeuProTyrThrIleEnd
1546  GAATTAAAATAGAAAGATGAAAGCGAATAAAAAAAAAGA
GAAAGCGCTTTACCCGGATGAACCTGCTCTGACAATACCACACAAAAATGGTGTGATA
S1606 ACAACCAACAAAA 1618
M  TACACTACCAATA
Oe  ATGCCGAAAGGTT
Figure 4.2

Sterile

Fertile

123
**Figure 4.3**
A Putative Transcript Processing Site in the 5' region of *atpA* in Sunflower

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sunflower atpA</strong></td>
<td>TTGTTCCTTTG-31-ATG</td>
</tr>
<tr>
<td><strong>Maize atpA</strong></td>
<td>TTTTTTCTTTT-45-ATG</td>
</tr>
<tr>
<td><strong>Oenothera atpA</strong></td>
<td>TCTTTTCTTTT-56-ATG</td>
</tr>
<tr>
<td><strong>Maize coxl</strong></td>
<td>GAAACTCTAGAATGA-86-CC TACATTTTTG-66-ATG</td>
</tr>
<tr>
<td><strong>Petunia atp9-1</strong></td>
<td>AAATTTCTTCAGATA-120-CCT-ATGCTTTG-124-ATG</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>AAATNTCATAAGTA (I)</td>
</tr>
<tr>
<td></td>
<td>TTNNTTTTTTG (II)</td>
</tr>
</tbody>
</table>

Sequences in the 5' flanking region of the *atpA* genes from sunflower, maize and *Oenothera* and the mapped 5' transcript termini of the maize *coxl* and *Petunia atp9-1* genes are aligned to show their similarity. For each gene the coding strand of mtDNA is shown with the 5' end to the left. The ATG at the right hand end indicates the initiation codon for each gene. The numbers between the sequences indicate the number of nucleotides that separate them. Asterisks below the *coxl* and *atp9-1* indicate sites at which the 5' termini of the RNA transcripts have been mapped. The shortest transcripts of *coxl* and *atp9* map to the sequences within the short consensus (II) sequence also found immediately upstream of the *atpA* genes. This sequence does not resemble the putative promoter consensus sequence (I) for plant mitochondrial genes to which longer transcripts of the *coxl* and *atp9* genes map (see section 1.8.3). It has been reported that transcripts are processed at the site defined by the consensus sequence II (Young et al. 1986). Nucleotides which match the consensus sequences are underlined.

No similarity to the proposed plant promoter consensus (I in Figure 4.3) was found in the 76bp of sequence obtained for the 5' region of the sunflower *atpA* gene. This was not unexpected as homology to this consensus in other plant mitochondrial gene sequences generally occurs at least 120bp upstream of the ATG (Young et al. 1986).

The sunflower *atpA* gene contains four GAA motifs immediately 5' to the ATG, similar to GA-rich short repeats found in the maize and *Oenothera* genes. Although these sequences show divergence between all three species, the
reiterative character is retained and this may be important in correct alignment of the 5' terminus of the transcripts during initiation of translation. A similar situation is found upstream of the chloroplast atpA gene from tobacco where there are two direct repeats of the sequence 5' AAAG 3'. In the chloroplast gene, these sequences are thought to represent ribosome binding sites as they are complementary to a sequence 5' CUUU 3' present at the 3' end of the chloroplast 16S rRNA (Deno et al. 1983).

As discussed earlier (section 1.8.6) the sequence of the 3' region of the mitochondrial 18S rRNA in wheat has been used to define complementary putative ribosome binding sites near the start of translation in plant mitochondrial genes (Bonen et al. 1984, Moon et al. 1985). Sequences immediately upstream of the plant mitochondrial coding regions such as the sequence 5' AGAATTCG 3' located 10-17 bases upstream of atp9 in Petunia are thought to mediate ribosome binding via complementary base pairing with the 3' TCCTAAGTT 5' sequence of 18S rRNA (Rothenberg and Hanson 1987). A slightly different sequence showing some complementarity to the wheat 18S rRNA 3' region is present 17bp upstream of the start of atpA gene in sunflower (5' CGATTGTA 3'), indicated by the dotted overline in Figure 4.2a) and may assume a similar function in ribosome binding. The repetitive GAA motifs immediately 3' to this putative binding site give the sunflower atpA gene an A-rich region between the putative binding site and initiator codon as has been reported for the maize coxl and cob genes (Isaac et al. 1985b, Dawson et al. 1984).

Within the coding region of the atpA genes from sunflower, maize and Oenothera, some nucleotides are conserved between only two of the three species and some are variable in all three. Non-conserved nucleotides are more prevalent in the 3' region. Of the 33 mismatches between sunflower and maize which predict amino acid changes, 21 are in the second half of the coding region. The atpA ORF of sunflower extends for 1530bp ending in a TAA 3' termination codon at +1531 (Figure 4.2a). The sunflower atpA open reading frame is 6 nucleotides longer than the maize coding sequence, which terminates at a TAG codon, and is 3 nucleotides shorter than the Oenothera sequence which terminates at a TGA codon. No long stretches of sequence homology are apparent in the 3' non-coding sequences of the three atpA genes (although see section 4.5).

4.4 Comparison of the Predicted Amino Acid Sequence of the Sunflower Mitochondrial atpA gene with Other F1ATPase α-Subunit Proteins

Translation of the ORF encoding the atpA gene from sunflower predicts a protein of 510 amino acids compared to 508 amino acids in maize and 511 amino acids in Oenothera. The three sequences are 92% homologous at the amino acid level and as indicated in Figure 4.4, the predicted amino acid sequence of the sunflower atpA gene shows a high degree of similarity to the
nuclearly encoded mitochondrial subunits of yeast (64%) and cow (68%, Walker et al. 1985) and also the corresponding protein in the tobacco chloroplast (55%, Deno et al. 1983) and bacterial ATPase complexes (49%, Gay and Walker 1981). Figure 4.4 shows that most of the homologous α-subunit polypeptides begin within two amino acids of the predicted sunflower initiator methionine. Both the nuclearly encoded bovine and yeast proteins are made as higher molecular weight precursor molecules, yet the mitochondrially encoded sunflower protein is only a single amino acid shorter than the corresponding yeast polypeptide and two amino acids longer than the mature bovine polypeptide. A similar observation with the maize atpA gene sequence, led to the proposal that, amino terminal leader sequences for nuclearly encoded mitochondrial proteins such as the yeast atpA, may have been 'acquired' independently from the protein coding region during the evolutionary transfer of genes from the 'original endosymbiont' to the nucleus (Isaac et al. 1985a).

In common with the yeast and bovine α–subunit polypeptide, the predicted sunflower sequence contains amino acids at positions 191-200, which are not present in the plastid or bacterial sequences (dashed line in Figure 4.4). The proteins in the four eukaryotes lack 11 amino acids which are found in the E.coli subunit, inserted at amino acid position 322. By comparing other adenine nucleotide binding proteins to both the α and β ATPase subunits of bovine mitochondria and E.coli, Walker et al. (1985) identified two domains ("A" and "B") which were highly conserved. The inference that these sequences are responsible for the binding of nucleotides is supported by their similarity in all α-subunits sequenced to date. The "A" domain is perfectly conserved between the proteins aligned in Figure 4.4 (also in the maize and Oenothera proteins, not shown) while the "B" domain is conserved in all but the tobacco chloroplast sequence which lacks the glycine residue corresponding to the sunflower amino acid residue at position 264 and which is found in all other adenine binding proteins (Walker et al. 1985, Deno et al. 1983).

The molecular weight of the atpA polypeptide from sunflower was calculated by computer translation and composition analysis of the gene sequence obtained. This predicts a value of 55.5 kilodaltons (kDa) for the sunflower polypeptide, which is approximately 3kDa lower than that estimated by SDS-polyacrylamide gel electrophoresis (see Chapter 6). An analysis of codon usage in the sunflower atpA gene indicates a bias for A in the third position (31% of codons end in A). A total of 56% of the codons contain A or T in the third position. This is slightly lower than the 'preference' for A or T in the third codon position which appears to be diagnostic of plant mitochondrial genes. In maize other reported values for A or T in the third position are 63% for coxl, 60% for coxII, 67% for cob and 66% for atp9 (Dewey et al. 1985).
Figure 4.4
Homology Between the Predicted Amino Acid Sequences of the $F_1ATPase \alpha$-subunit Proteins from Different Organisms.

This figure aligns the predicted amino acid sequences (in single letter code) of the sunflower (S) ATPA polypeptide with the corresponding proteins from *Saccharomyces cerevisiae* (Y), *Bos taurus* (bovine B), *Nicotiana tabacum* chloroplast (T) and *Escherichia coli* (E). Numbers refer to the sunflower sequence and only this sequence is shown in full. The amino acids are shown in the other sequences only where they differ from the sunflower sequence. The residues underlined between 171-178 and 261-271 are the respective "A" and "B" ATP nucleotide binding domains described by Walker *et al.* (1985). + in the *E.coli* sequence marks the position (320) where 11 amino acids are inserted relative to the other sequences (AFTKGEVKGKT, not shown in the amino acid sequence alignment).
4.5 Sequence Divergence Occurs 3' to the atpA Genes in Fertile and Sterile Sunflower

Comparison of the sequences present downstream of the atpA genes from fertile and sterile sunflower revealed that a base mismatch is present 161bp downstream of the end of the coding region (nucleotide position +1691) where an A in the sequence from fertile sunflower is replaced by a G in the sequence from sterile sunflower (shown by the gels in Figure 4.2c). A further 37 nucleotides downstream of this mismatch the sequence in the sterile sunflower completely diverges from that of fertile (nucleotide position +1728). The sequence of the 3' flanking region of the atpA gene, covering the regions of divergence between fertile and sterile sunflower are shown in Figure 4.5a. The 40 carboxy-terminal amino acids of the atpA ORF (orfA) are including to indicate where homology between the 3' non-coding sequences of the two atpA genes ends, relative to the coding region.

In fertile sunflower, at a distance of 320 nucleotides downstream of the atpA termination codon there is an inverted repeat sequence (indicated by the inverted arrows begining at position +1851 in Figure 4.5a). This region has the potential to form, by complementary base-pairing, the stem-loop secondary structure shown in Figure 4.6a (page 137). This secondary structure is similar to the stem-loops found in the 3' non-coding region of the atp9 genes of Petunia and tobacco which are thought to function in transcript termination (Rothenberg and Hanson 1987, Bland et al. 1986). A different structure, proposed as the putative terminator of atpA transcription in Oenothera is shown in Figure 4.6b. In Oenothera, the atpA mRNA is reported to be 2kb, initiating approximately 250bp upstream of the ATG codon and terminating 265 downstream of the termination codon (+/-3 nucleotides, marked by vertical arrowheads in Figure 4.6b) (Schuster et al. 1986).

Immediately after the predicted stem-loop structure in the 3' region of the atpA gene from fertile sunflower, there are 10 nucleotides very similar to the sequence which includes the termination sites for atpA in Oenothera (underlined in 4.6a). This may represent the point of atpA transcription termination in fertile sunflower and an alignment of this region with sequences surrounding mapped 3' transcript termini of the Oenothera coxl and maize coxl genes is shown in Figure 4.6c. A sequence analogous to this putative transcription termination site is not found in a similar position in the sequence of the atpA gene from the sterile line.
Figure 4.5a
Sequence Divergence Between the 3' Flanking Region of atpA in Fertile and Sterile Sunflower

The DNA sequence of the 3' coding and flanking non-coding region of the atpA gene contained within the 2.2kb EcoRI from fertile (F) and 3.8kb EcoRI fragment from sterile (S) sunflower are shown. The sequence shown begins at position +1411, relative to the A (+1) of the ATG start codon (compare Figure 4.2a) and the predicted carboxy-terminal amino acid residues (in the single letter code), which are identical in the two genes, are also included. The 1.2kb EcoRI fragment from sterile sunflower which hybridises with the 3' region of the atpA gene (i.e. clone pA212) was found to contain 739 nucleotides of sequence homology to the flanking region of atpA from the fertile genome. The extent of sequence identity is indicated by the vertical lines between this 1.2kb EcoRI fragment (pA) with the atpA 3' sequence from fertile sunflower.

The open reading frames (ORFs) present within the atpA 3' flanking sequence from fertile sunflower (orfb and orf70, coding strand as atpA) and from sterile sunflower (orfC, coding strand as atpA, and orfd, complementary strand) are indicated by arrows at the initiating methionines and asterisks at the termination codons. The predicted amino acid sequences for these ORFs are included. Single base mismatches (point mismatches) are denoted by an X between bases, and a vertical arrowhead. An inverted repeat sequence 320 bases downstream of the atpA gene in fertile sunflower is indicated by inverted arrows.
Figure 4.5a (continued)

3331S ATTCATCCTCTCCTATGCTCTTCCAGCGCGTGTGTAGCTTTTCTCAATTCCTCTCGCTA 3390

3391S TCGTTAAGGCAACTATGTAACCTTTTCTCTCGCTACTGTCGCTTTCTCTCGTACTGTA 3450

3451S GCAGACAAGCGAGCTATATGGGCTGGAGGAGTGTCTGAGTATACATCG (orf ends) * T G K Q R Y Q T K R K Y T C R 3510

3511S CGTGGTAAAGCAACTATGTACCTTTCTGTCGGTACTGTGTCTTTCGCTTGTAGGTACATCG 3570

3571S GATACATGGAAAGACAGAAATGACACAGAAAGCGAACTACCATGTAGC 3630

3631S (orfd ends) T G K Q R Y Q T K R Y T C R

3691S TAGACTCTTTTCTCTCCTATGTTACCGCTCATGTTCGGCTGGGAGCTGGGCTTGCTTCGCTAATGGATATATAGGTATGGGAA 3750

3751S GAGGAAATCGGAGGGTGGGAAAGTCTGTTACTTAAG 3' (S) L F D T G A L R D F L K F L C K L S F C

D G L V G V K L C H I? <- orfd
The Region of Sequence Divergence 3' to the Sunflower *atpA* Gene Shows Homology to Other Plant Mitochondrial DNA Sequences

Sequence homology between the region 3' to the *atpA* genes from fertile (F) and sterile (S) sunflower and the 5' region of the tobacco *atp6* gene (Nt5') and *Oenothera coxIII* (Oe5') gene is shown (Bland *et al.* 1986, Hiesel *et al.* 1987). Asterisks denote nucleotide matches between sunflower and tobacco (above the Nt5' sequence) and between sunflower and *Oenothera* (below the Oe5' sequence). The numbering of the sunflower *atpA* 3' sequences are as Figure 4.2a (the TAA termination codons at position +1531 are underlined). The numbers for Nt5' and Oe5' are relative to the ATG (+1) initiation codons of the *atp6* and *coxIII* genes respectively. The ATG codons for *atp6* in tobacco, orfb and orfc in sunflower and an orfB in *Oenothera* are shown in bold typeface. Beyond the point of sequence divergence between fertile and sterile sunflower (closed triangle) homology between the *Oenothera* orfB and the orfb sequence of fertile sunflower is indicated by asterisks, and homology between orfb and orfc sequence of sterile sunflower is indicated by vertical lines. A possible ribosome binding site, immediately upstream of the ATGs in all four sequences is denoted by dots (see c). The TGA codons for orfs b and c and the TAA codon for the *Oenothera* ORF are underlined (note the entire orfB sequence is not included).

A Potential Ribosome Binding Site Associated with the Open Reading Frames orfb and orfc in Fertile and Sterile Sunflower

Potential base pairing between the region upstream of the ATG codons for orfb and orfc in sunflower (dots in b) and the 3' end of the 18S rRNA from wheat (Chao *et al.* 1984) is indicated by vertical lines. The part of the 18S rRNA equivalent to the Shine-Dalgarno sequence from *E.coli* is underlined (Shine and Dalgarno 1974).
Figure 4.5b

+1525  F  ACAATATTTAAAAGAAGAAAGATAAAAATAGAAAGATGAAGGAACAAAGTTGACACAATCCCTT
  S  ACAATATTTAAAAGAAGAAAGATAAAAATAGAAAGATGAAGGAACAAAGTTGACACAATCCCTT
  ****  ***  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
-173  Nt5'  TATATATGAAATTACTTCGTCCTTTTTTTTAGCCCTTTTTCGTTTGTCCATCTTTTTTTCTCCCAT
-179  Oe5'  GTGCCCAAGGCAATGTTACTACGCGCAGGCGGCAAGGTTGTTGACATCAAGGCTTTTTTTCCTGATG
  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

1591  F  TCTTTCGTTGTCACAACAAACAAAGAAATCTGTTT--AGT-----TCTTCACCTACTGCGATACAG
  S  TCTTTCGTTGTCACAACAAACAAAGAAATCTGTTT--AGT-----TCTTCACCTACTGCGATACAG
  *******  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

1651  F  GAAGGCTCTCTCTTT--CTGTTGCCCCGGAATCTTTATTCTCAATTAACAG--ATGCGCTCAACTGGA
  S  GAAGGCTCTCTCTTT--CTGTTGCCCCGGAATCTTTATTCTCAATTAACAG--ATGCGCTCAACTGGA
  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

1712  F  TAAATTCACTTTATTTTCGCTTTACTATTTTGCTATTTTTCACATTTGCTACTCGGACGTTTTCT
  S  TAAATTCACTTTATTTTCGCTTTACTATTTTGCTATTTTTCACATTTGCTACTCGGACGTTTTCT
  *********  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

1778  F  AGGAACGGGCTGCCCCGGGATTTCCACCCTATTGACTTTATTTTTTATATAG
  S  AGGAACGGGCTGCCCCGGGATTTCCACCCTATTGACTTTATTTTTTATATAG
  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

1844  S  ATTTCTGCCATATCGCAAAGCATCTCCCCTCCTGGTGGATCCGGCGATCGACCAATAGAAGAAT
  Oe5'  GGTAGAACATCTCCGCGACACACCCAACAGCGTTGGAGAAGACTCTTGTAAAGATGTTTGACACC
  I  II  I

1910  S  CCGTGGTCTGCTTAAAGGACATCTGCCGCGCGGAGTTTACTCCGGCGCACTGGTTCTCTGTA
  Oe5'  GGTGATATCTCTTATATGTATCTACTGTTATTCTGCGATCTCCAAAATGGTGTAAGGCTGTCG>183TGA
  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I  I

Figure 4.5c

orf b/orf c 5'  5'  GGGGGAAUCUCCCUGUA  3'  3'
18S rRNA  3'  CUUCCUAAGUAGUAG  5'
The 3' Flanking Region of the *atpA* gene in Fertile Sunflower has the Potential to Form a Stem-loop Structure and Shows Homology to Transcript Termination Sites in Other Mitochondrial Genes

The stem-loop secondary structure which can be formed from the 27 bp inverted repeat sequence located 320bp downstream of the TAA termination codon of *atpA* in fertile sunflower is shown in (a). Double inverted repeats in the 3' flanking region of the *atpA* gene in *Oenothera* predict the formation of the structure shown in (b). A region in the sunflower *atpA* 3' flanking region which is similar to the sequence containing three *atpA* transcript termination sites in the *Oenothera* sequence (denoted by the vertical arrows in b) is underlined in (a). Sequence homology between the predicted transcript termination region of the *atpA* genes from sunflower and *Oenothera* with *coxII* from *Oenothera* and *coxI* from maize are aligned in (c).
Figure 4.6

(a) Sunflower (fertile) atpA

(b) Oenothera atpA

(c) Species | Gene | Distance from ATG | sequence around predicted 3' transcript termination site
--- | --- | --- | ---
Sunflower atpA | 1911 | ATCTACGTGT - -TATGTACTT
Oenothera atpA | 1877 | TTCACGTGTGCTATGATAGG
Oenothera coxII | 1089 | TTCACGT - - TATGATAGG
maize coxl | 2045 | TTCATGGGT - - T - CGATGGC
4.6 Open Reading Frames Downstream of the *atpA* gene in Fertile and Sterile Sunflower

An open reading frame of 210 nucleotides, potentially encoding a polypeptide of 70 amino acids (designated orf70) is present downstream of the *atpA* gene (orfA) in fertile sunflower (indicated in Figure 4.5a with the methionine initiation codon corresponding to orf70 at position +1953). The orf70 sequence shows a bias for A or T in the third codon position of 58.5% in line with other mitochondrial coding sequences. No significant homology to orf70 was found by searching the GenBank and EMBL computer data systems (Devereux *et al.* 1984). Orf70 is not present in the region downstream of *atpA* in the 3.8kb *EcoRI* fragment from sterile sunflower.

A second ORF of 132 nucleotides, (designated orfb) is located between *atpA* and orf70 in the sequence from fertile sunflower beginning at nucleotide position +1698 (Figure 4.5a). This ORF includes the point at which the sequence divergence occurs in sterile sunflower and a corresponding ORF, in the sterile sunflower, shares the first 30 nucleotides with orfb, then extends for a further 243 nucleotides (this extended version is designated orfc). The sequences of orfb and orfc potentially encode polypeptides of 43 and 90 amino acids respectively, including an identical first ten amino-terminal residues.

Initial computer searches using the the orfb and orfc sequences proved inconclusive. The region which includes the *atpA* associated sequence divergence between fertile and sterile sunflower was previously shown to hybridise to several bands in restriction endonuclease digests of sunflower mtDNA, indicating it to represent some form of multiply repeated element within the genomes (see Figure 3.24). Reports in several other plant species suggest that some short sequences, found in the flanking regions of known mitochondrial genes, are duplicated or more extensively repeated in plant mtDNA and may be conserved between different species (Hiesel *et al.* 1986, Bland *et al.* 1987, Folkerts and Hanson 1989). An extensive analysis of published plant mtDNA sequences, including those from non-coding regions of various mitochondrial genomes, led to the finding of a significant level of homology between the 3' region of the *atpA* genes in fertile and sterile sunflower and sequences from both *Oenothera* and tobacco mtDNA. Of particular interest, was the fact that the homologies found were associated with mtDNA sequences which are part of repeats within these other genomes.

Within the 3' non-coding region of the sunflower *atpA* gene, before the point where the sequence in the genomes from fertile and sterile diverge (closed triangle in Figure 4.5b) there is homology to a 657bp repeat found in *Oenothera* mtDNA upstream of the genes for *coxl* and *coxIII* (Hiesel *et al.* 1987). A sequence showing some similarity to the *Oenothera* 657bp repeat is also repeated in tobacco mtDNA, upstream of *atp6* and in the intergenic region.
of atp9 and rps13 (Bland et al. 1987). The extent of homology between these other plant mtDNA sequences and the sequenced region downstream of the sunflower atpA genes is shown in Figure 4.5b. The TAA termination codon of the sunflower atpA genes is included at the start of the alignment shown (compare Figures 4.2a and 4.5a) and indicates that good homology with the Oenothera repeat (Oe 5') begins 27 nucleotides downstream of the end of the atpA coding region. A further 28 nucleotides downstream, homology with the sunflower atpA 3' flanking sequences and the tobacco atp6 5' region (Nt 5') begins abruptly and continues for approximately 160bp (the sequence for the 5' region of the rps13 gene in tobacco is identical to that shown for the atp6 gene and so is not included). The homology between the sunflower and tobacco genomes, consists of three almost perfectly conserved 'domains' of 34bp (positions +1584 to +1618 in the sunflower atpA 3' sequences), 18bp (+1632 to +1651) and 32bp (+1665 to +1696). The 34bp domain is partially conserved in the Oenothera repeat and both the 19bp and 32bp domains are highly homologous in all three species. The border of the 657bp repeat of the Oenothera genome actually falls within the 32bp domain, at the T residue in the sequence CAAT, corresponding to position +1693 in the sunflower sequences.

Homology between the sunflower and tobacco sequences ends abruptly at the ATG codons corresponding to the start of the atp6 (or rps13) gene. The methionine initiation codons for orfb in fertile sunflower, orfc in sterile sunflower and an unassigned ORF (designated orfB by Hiesel et al. 1987) in the Oenothera sequence occur two nucleotides upstream of the position corresponding to the tobacco atp6 ATG. Homology between the fertile and sterile sunflower sequences and the Oenothera repeat continues into the open reading frames of orfb, orfc and orfB respectively. The orfB sequence in Oenothera is 477bp, potentially encoding 159 amino acids. The first 10 codons of orfb in fertile sunflower, and the first 19 of the orfc in sterile sunflower are completely identical with those of the orfB in Oenothera. The extended homology between the sterile sunflower orfc and the Oenothera orfB includes the region where orfc and orfb sequences diverge (closed triangle). Indeed, the 25 nucleotides beginning at the point of divergence in the sterile sunflower sequence are completely identical in Oenothera (vertical lines immediately after the closed triangle in Figure 4.5b). Beyond this region, the level of homology with the orfb, orfc and orfB is less than 20%. The sequence downstream of the TGA codon of orfc shows no significant homology with the rest of orfB of Oenothera (the rest of the orfB sequence is not shown).

As stated, the 3' border of the 657bp repeat from Oenothera occurs 5 nucleotides upstream of the ATG of the orfB sequence in this genome which is 4 nucleotides upstream of the initiation codons for orfb and orfc in fertile and sterile sunflower respectively. Thus, as part of the single-copy flanking region
of the 657bp repeat, orfB is unique to the two coxIII 'alleles' which occur in Oenothera through homologous recombination across the 657bp repeat (Hiesel et al. 1987). The end of the orfB sequence is over 600 nucleotides upstream of the actual coxIII ATG codon. The mRNA transcript for coxIII is initiated from a sequence 5'GTAAGTGA 3' which occurs in a central position in the 657bp repeat, a distance of 340 upstream of the start of the orfB sequence. The region between the 5' terminus of the coxIII mRNA and the 3' border of the 657bp repeat therefore contain the promoter for initiation of coxIII transcription (Hiesel et al. 1987). This putative promoter is beyond the region where homology with the Oenothera repeat and the sunflower atpA 3' sequence begins.

Part of the conserved 32bp domain in the sunflower, tobacco and Oenothera sequences includes a potential ribosome binding site (dots in Figure 4.5b) which is shown aligned with the 3' sequence of the wheat 18S rRNA in Figure 4.5c. Thus, this region of sequence divergence between fertile and sterile sunflower contains features in common with the 5' region of functional genes in the tobacco mitochondrial genome (i.e. atp6 and rps13) and Oenothera (coxI and coxIII) mtDNA. The putative ribosome site is additional evidence that the orfb sequence encoded by this region in fertile sunflower and the modified version, orfc, in sterile sunflower may be expressed.

The exact function of the conserved sequence homology found in the sunflower, tobacco and Oenothera mitochondrial genomes sequence is unknown. Apparently, other regions of homology exist in the sunflower mitochondrial genome (Figure 3.24) and similar findings in other species supports the proposal that this conserved sequence may have a role in the creation of genome reorganisation. Bland et al. (1987) have reported that, in addition to the 5' flanking sequence of the atp6 and rps13 genes, at least four other copies of sequences homologous with these regions are present in tobacco mtDNA. More recently, a probe from the intergenic region of the tobacco atp9-rps13 genes was found to hybridise to eight different DNA restriction fragments from various regions of the Petunia mitochondrial genome, none of which correspond to the atp9-rps13 intergenic region in this genome (Folkerts and Hanson 1989). The significance of the findings in sunflower are further discussed in Chapter Seven.

Analysis of the sequence further downstream in the region of divergence between fertile and sterile sunflower indicated that the DNA strand complementary to the one encoding atpA contains a large ORF in the genome from sterile sunflower. This ORF (designated orfd shown in Figure 4.5a) extends for 382 nucleotides, (corresponding to 127 amino acids) in a 5' to 3' direction from the opposite end of the 3.8kb EcoRI fragment which contains the atpA gene in this genome. An analysis of codon usage within this ORF indicates a slight bias for A or T in the third position (54.7%). However, the
only methionine codon occurs as the ninety-third codon in the 127 codons of sequence obtained for orfd. If this ORF is translated, the initiator methionine is probably located in the fragment adjacent to the 3.8kb EcoRI, i.e. the 2.4kb EcoRI located during the mapping studies (see Figure 3.23). Orfd lies within the region of the sterile sunflower mtDNA which the mapping and hybridisations studies suggest arose by a sequence insertion. Translation of the nucleotide sequence of orfd predicts a protein of molecular weight greater than 14.4kDa but searches of the databank revealed no significant homologies to any other proteins.

4.7 Sequence Analysis of the Inversion Endpoint in the Mitochondrial Genome of Sterile Sunflower

The 1.2kb EcoRI fragment from sterile sunflower mtDNA, which maps to the end of the region of sequence inversion described in Chapter three, was partially sequenced by myself and also by Dr. Ray Vonder Haar, Texas A&M University. This fragment, was subcloned from the plasmid clone pA212 into mp19, and sequenced from both ends and by 'oligopriming' across the central region. The fragment was found to contain 739bp of sequence homologous to the region from positions +1463 to +2202 in the 2.2kb EcoRI fragment from fertile sunflower as shown in Figure 4.5a (where the homologous sequence within the 1.2kb EcoRI fragment is designated pA). This region of homology therefore includes 68 nucleotides, corresponding to the last 22 amino acids of the carboxy-terminus of the atpA coding sequence, plus the 671 nucleotides from the TAA termination codon to the end of the 2.2kb EcoRI fragment (and hereafter will referred to as the 'pA repeat'.

The pA repeat spans the region of divergence between fertile and sterile sunflower, in the 3' non-coding sequence of the atpA gene, and thus a homologous 'core' of 266bp, corresponding to position +1463 to +1728 is common to all three sequences. Approximately 100bp of this 'core' sequence form the region of homology found with the Oenothera and tobacco mitochondrial genomes described above. This leads to the conclusion that homologous recombination within this core sequence in the genome of fertile sunflower, probably caused the inversion in the mtDNA from sterile sunflower and the relocation of the atpA 3' flanking region approximately 13kb downstream thus creating the 1.2kb EcoRI fragment. The remaining 480bp of sequence of the 1.2kb EcoRI fragment contains no homology to the atpA flanking region and is shown in Figure 4.7a.
Figure 4.7
An Open Reading Frame is Located Adjacent to the Sequence Inversion in Sunflower Mitochondrial DNA
Sequence for the part of the 1.2kb EcoRI fragment from sterile sunflower beyond the homology to the atpA 3' region is shown in (a). The numbering begins at the EcoRI site downstream of the end of the inversion in this genome and corresponds to the sequence presented in Figure 4.5 (pA) beginning at nucleotide position 429. The predicted amino acid sequence of an ORF (orf73) partly encoded by the 1.2kb EcoRI is indicated below the corresponding nucleotide sequence. A summary of the 'pA repeat' and 'single copy' regions of the 1.2kb EcoRI fragment is shown below the sequence. The diagonal shading indicates the atpA coding region homology (5' to the right for this coding strand) within the repeat. The orientation of orf73 is indicated by the direction of the arrow.

(b) The cosmid clone A5C3 which contains the downstream region of the mtDNA inversion from sterile sunflower was digested with EcoRI (lanes 1 and 3), AvaI (lanes 2, 4 and 5) or both (lane 6). The 4.2kb EcoRI fragment from this clone, situated adjacent to the 1.2kb EcoRI (see Figure 3.23) was labeled with $^{32}$P and hybridised with lanes 1 and 2. The parental 4.2kb EcoRI (in lane 1) hybridises to a ca. 900bp AvaI fragment plus the remainder of the 4.2kb (3.7kb lane 2). When the 1.2kb EcoRI fragment is used as a probe (lane 3), the same 900bp AvaI fragment hybridises (lanes 4 and 5). The region of the 900bp AvaI which hybridises can be localised to the 410bp AvaI-EcoRI fragment (lane 6) from the 1.2kb EcoRI. As the 410bp AvaI-EcoRI includes the repeated region of the 1.2kb EcoRI fragment this confirms the orientation of the repeat to be indirect with respect to the intact copy of the atpA gene present in the genome of sterile sunflower.
Figure 4.7b

Figure 4.7a

5' 1 GAATTCCTAATTTTACACAGAAAATGAAAACAGAAATCACATGATTACCAG
  EFLIFTQKMKTNQMITTS
  51 TAGACAATCTGATCTACGAGGATGGTACGCTCTCAATGAGACGG
  RQLVLRSEMDALMET
  101 ATTTGGATTTCTTCTTCTCCGGGATTCGCTGCTTCTCGCGGTT
  DWDSPSSADSAASSAY
  151 CGCTGCTCTCTCAACAGGTCGGCGACGGCGAGGCGTCCACTTGA
  CNYTKK*
  201 TTGCAATTATACAAAGAATAGTGATGATGAGGCGACGGCTTTTGGTTGA
  AluI
  251 AGCCGGGTATGTTATACACCAGGGCTTGACCATTGCGAGCTATTACGA
  301 GAACCTTGCGGATGACGGGTTGGGTTATTATATGGGAATACATTATT
  351 CATTTPAAAGCGGCGCTTTCTAGGGAAACCTACTAGGATTCTGATAAA
  AluI
  401 TAGCAATTGATAGTTTGAATTCCAGAAGCTACTGATTTTGCGCATGACT
  AluI
  451 TCAGCAAGTACGGTGAGTTTTAGTACGGCTGGAACACGAAAAGGAAGAAA 3'

pA repeat

orf73
The sequence of the 1.2kb EcoRI fragment indicated a single AvaI site (CCCGGG) was located within the pA repeat at a distance of 410bp from one EcoRI site (underlined in Figure 4.5a). This fragment was isolated from the clone pA212, labelled with $^{32}$P and hybridised with a cosmid containing the 1.2kb EcoRI and adjacent fragments (A5C3) which had been digested with AvaI, EcoRI or both enzymes. Figure 4.7b shows that the 410bp AvaI-EcoRI is part of a 900bp AvaI fragment spanning the 1.2kb EcoRI fragment and the adjacent to the 4.2kb EcoRI fragment (see Figure 3.23). This confirms that the orientation of the 1.2kb EcoRI fragment within the genome of sterile sunflower is such that the sequence homologous to the 3' region of the atpA gene which it contains is indirect with respect to the intact atpA gene. It was not possible to confirm the orientation of the pA repeat until the complete sequence of the 1.2kb EcoRI fragment was obtained, due to a lack of suitable sites for a range of restriction enzymes (ClaI, XhoI, PstI, BamHI, BglII, SacI, KpnI and HindIII).

The sequence the 1.2kb EcoRI fragment from sterile sunflower indicates that the pA repeat which it contains is not part of any larger ORF sequence as there is a lack of an initiator methionine in its upstream region (i.e. from positions 1 to 481 within the 1.2kb EcoRI sequence shown in Figure 4.7a). Within the 1.2kb EcoRI fragment the longest ORF occurs in the region beyond the repeat, and extends for 219 nucleotides from the opposite end of the fragment before terminating in a TAG codon at position 220 (Figure 4.7a). These 219 nucleotides potentially encode 73 amino acids (the predicted amino acid sequence is included in Figure 4.7a) hence the designation of this ORF as orf73. Four methionine codons occur within the sequence obtained for orf73, although it may form part of a longer ORF which begins in the contiguous 3.4kb EcoRI fragment (see Figure 3.23) which was not sequenced.

As part of the single copy region beyond the mtDNA rearrangement in sterile sunflower, the orf73 sequence should also be present in the 0.8kb EcoRI fragment from fertile sunflower which is homologous to the 1.2kb EcoRI from sterile sunflower. None of the mitochondrial genes probes which were used for the mtDNA hybridisation analysis described in Chapter three, other than that for the atpA 3' region, were found to show homology to the 1.2kb EcoRI fragment from sterile sunflower suggesting that, if the orf73 coding sequence partly encoded within this fragment is indeed expressed, it does not encode any of the genes previously determined to be encoded by the sunflower mitochondrial genome.
4.8 DNA Sequences Involved in Generating the Rearranged \textit{atpA} locus in the Mitochondrial Genome of Sterile Sunflower

The sequence analysis indicates that the 1.2kb \textit{EcoRI} fragment containing the pA repeat in mtDNA from sterile sunflower represents one product of intragenomic recombination, and has the 3' end of the 2.2kb \textit{EcoRI} 'progenitor' fragment present in fertile sunflower. As the other product of recombination, the 3.8kb \textit{EcoRI} \textit{atpA} containing fragment in sterile sunflower, would be expected to contain part of the 0.8kb \textit{EcoRI} fragment implicated as the other 'progenitor' sequence in the fertile sunflower. Unfortunately the sequence of this 0.8kb \textit{EcoRI} fragment was not obtained, thus precluding a precise definition of the recombinational breakpoints involved in the inversion. However, several features of the putative recombination events which have created the rearrangement associated with \textit{atpA} in sterile sunflower mtDNA can be proposed.

A combination of the hybridisation and sequence analysis suggests that the 0.8kb \textit{EcoRI} fragment from fertile sunflower is likely to contain at least the 266bp 'core repeat' found to be common to the 2.2kb \textit{EcoRI} from fertile sunflower and the 3.8kb and 1.2kb \textit{EcoRI} fragments from sterile sunflower. If this is the case, then the recombination which has resulted in the inversion in sterile sunflower will map to a point between the boundaries of this common sequence. As described earlier (section 4.5), there is a point mismatch at position +1691 between the sequences of the 2.2kb \textit{EcoRI} fragment (an A nucleotide) and the 3.8kb \textit{EcoRI} fragment (a G nucleotide), marked by an X between the nucleotides at an arrowhead above the position in Figure 4.5a). In the repeat present on the 1.2kb \textit{EcoRI}, the corresponding position is occupied by an A nucleotide (position 708 in the pA sequence in Figure 4.5a). Further upstream, a T nucleotide at position +1466 in both the 2.2kb and 3.8kb \textit{EcoRI} is mismatched by a C in the pA repeat (position 484). If the 0.8kb \textit{EcoRI} fragment contains G residues corresponding to both these positions, this could indicate the point of recombination was located between the two mismatches as shown:

\[
\begin{align*}
\text{Mismatch:} & +1466 & +1691 & \text{+1466} & +1691 \\
2.2kb & 5' \text{T} & \text{A} & 3' & 5' \text{T} & \text{G} & 3' & 3.8kb \\
0.8kb & 5' \text{G} & \text{G} & 3' & 5' \text{G} & \text{A} & 3' & 1.2kb
\end{align*}
\]
However, simple point mutations may have generated these differences and it must be noted that there are two other mismatches between the sequence of the 2.2kb EcoRI fragment and the 1.2kb EcoRI, which are located 3' to the proposed region of recombination and where they would be expected to be identical (also marked by Xs and arrowheads in Figure 4.5a, C at +1960 in the 2.2kb EcoRI and an A at the corresponding 978 nucleotide position in the 1.2kb EcoRI fragment, also a C at position +2041 in the 2.2kb EcoRI fragment and a G in the corresponding 1059 position in the 1.2kb EcoRI fragment). Interestingly, the mismatch at position 1059 within the 1.2kb EcoRI fragment is part of the orf70 reading frame (described in section 4.6) and changes a predicted TCA codon (serine) in the 2.2kb EcoRI sequence from fertile sunflower to a TGA (codon stop) in this 'repeat' within the sterile cytoplasm.

Several conclusions, concerning the mechanism by which the rearrangement in sterile sunflower arose, can be made on the basis of the sequence data described:

1. The 739bp of sequence homologous to the 3' flanking region of atpA present on the 1.2kb EcoRI fragment of sterile sunflower has been derived, in part, from the 0.8kb EcoRI fragment which was the probable progenitor repeat containing fragment in fertile sunflower genome. The rest of the 739bp homology has been derived from the region 3' to the atpA gene contained within the 2.2kb EcoRI of fertile sunflower, with the point of recombination being at the junction of the two sequences.

2. Within the 3.8kb EcoRI fragment from sterile sunflower, there will be a sequence derived from the 0.8kb EcoRI fragment in the region downstream of the recombination site. This may be represented by the sequence beginning at the point of divergence between the 2.2kb EcoRI and 3.8kb EcoRI fragments, which would therefore constitute one of the unique borders of the progenitor repeat. If this were the case, the size of the repeat within the 0.8kb EcoRI mtDNA fragment of fertile sunflower must be less than 250bp (i.e. can span, at the most, the 229bp corresponding to positions +1462 to +1691 in the atpA flanking region). This would explain why it was not accounted for on the published physical map of the sunflower mitochondrial genome (Siculella and Palmer 1988).

3. Recombination between the 0.8kb and 2.2kb EcoRI fragments within the fertile sunflower genome does not entirely account for the organisation of the 3' end of the 3.8kb EcoRI fragment in sterile sunflower. This is due to a sequence insertion, which encompasses the end of the 3.8kb EcoRI fragment and at least 400 bases of the adjacent 2.4kb EcoRI fragment (see Figure 3.26) a distance of approximately 2.3kb.

Sequence analysis of the 0.8kb EcoRI fragment from fertile sunflower would indicate the extent of homology between this and the 3.8kb EcoRI fragment in the sterile genome and define which region within the 3.8kb
EcoRI has resulted from sequence insertion. However, the sequence insertion may also have arisen by homologous recombination which could complicate a definition of the repeats involved in the inversion as opposed to the insertion. The insertion could also be the product of more than one recombination event, involving different regions of the mitochondrial genome so that a multiply rearranged sequence is now present downstream of the atpA gene in sterile sunflower.

4.9 Conclusions

The DNA sequence data presented in this chapter shows that the atpA gene in sterile sunflower encodes a protein product identical to the one encoded by the mitochondrial genome in the fertile line. Recombination between the 3' region of the gene and an indirectly repeated sequence has displaced the original 3' end of the atpA gene and its flanking region to the opposite end of a ca. 13kb inversion in the genome of sterile sunflower. The intragenomic reorganisation has probably occurred via homologous recombination within a region of 229 nucleotides which is related to sequences known to be recombinogenic in other mitochondrial genomes. In addition, the 3' region of the atpA gene in sterile sunflower has been the site of insertion of approximately 2.3kb of sequence into this genome.

These recombination events do not affect the atpA coding sequence which is present in the mitochondrial genome of sterile sunflower and hence would not be expected to alter the protein product of this gene. However, a sequence which may act as a site of atpA transcript termination or processing in fertile sunflower has been removed from downstream of the atpA gene in sterile sunflower. The indirect repeat, which contains the atpA 3' coding region and associated flanking sequences located ca. 13kb from the intact gene in the genome from sterile sunflower, does not appear to be part of any ORF created by the rearrangement. Part of another ORF (orf 73), was found to map adjacent to the atpA repeated sequence within the 1.2kb EcoRI fragment from sterile sunflower although its identity remains to be shown. As demonstrated in the next chapter, this orf73 does appear to be expressed as part of a large transcript, in mitochondria from both fertile and sterile sunflower.

As a consequence of rearrangement downstream of the atpA gene, a small ORF of 43 codons in fertile sunflower (orf8) has been extended to over twice this length in sterile sunflower (orf8, 90 codons). Further downstream within the region of divergence an ORF corresponding to 70 codons (orf70) occurs in fertile sunflower. This orf70, which is present as part of the repeat on the 1.2kb EcoRI fragment in the sterile genome, is truncated due to a point mismatch which introduces a premature TGA stop codon after 29 residues.

The region thought to have arisen by sequence insertion downstream of atpA in sterile sunflower contains an ORF potentially encoding at least 127
amino acids (orf). Thus, these regions of the mitochondrial genome which have been reorganised in sterile sunflower contain ORFs which may be expressed. The consequences of these rearrangements for genome expression were investigated by an analysis of mitochondrial transcription patterns of fertile and sterile sunflower as described in the following chapter.
CHAPTER FIVE

TRANSCRIPTION OF THE MITOCHONDRIAL GENOME IN FERTILE AND STERILE SUNFLOWER

5.1 Introduction and aims

In order to investigate whether the differences in mitochondrial genome organisation between fertile and sterile sunflower are reflected in altered patterns of transcription, several mitochondrial gene probes were hybridised with sunflower mtRNA. The aims of the experiments described in sections 5.2-5.4 of this chapter were to:

a. Examine the possibility that the genome rearrangements 3' to the \(\text{atpA}\) coding region has consequences for the transcription of this gene in the sterile line;
b. Examine whether the \(\text{cob}\) gene transcription is affected by its linkage with the rearranged region of the genome;
c. Compare the transcripts homologous to other gene coding regions for evidence of variation between the cytoplasmic types.

In addition, a preliminary investigation of mitochondrial transcription during the reproductive phase of sunflower development was carried out. Vegetative development in male sterile sunflower is indistinguishable from that in fertile plants. The CMS phenotype is only expressed during microsporogenesis suggesting that any lesion in mitochondrial biogenesis and function, which causes the abnormal pollen development, does not become detrimental until differentiation of the male reproductive tissue. As discussed earlier, descriptions of microsporogenesis in sunflower, based on electron and light microscopic studies, point to changes in the tapetal cell layer as the earliest observable divergence from normal differentiation in anthers from CMS plants (Horner 1977). Normal pollen development in many species is characterised by intense metabolic activity in the cytoplasm during the phase of pollen maturation. Changes in mitochondrial gene expression leading to, and during, pollen development may be required to meet an increased demand for respiratory activity associated with this stage of development.

The specificity of the expression of the CMS phenotype in sunflower was addressed by analysis of RNA from several stages of flower development. Observations which were made from a comparison of the levels of total RNA from flowers of fertile and sterile sunflower and the patterns of expression of specific mitochondrial genes are described in sections 5.5-5.9.
5.2 *atpA* Transcription is Altered in Sterile Sunflower

To investigate *atpA* transcription, mtRNA was extracted from etiolated seedlings of the fertile, sterile, hybrid and restorer sunflower lines, fractionated by denaturing agarose gel electrophoresis, transferred to Hybond-N membrane and hybridised with specific gene probes ('Northern' hybridisation). The mtRNA was hybridised with the 2.2kb *EcoRI* fragment containing the *atpA* gene from the fertile cytoplasm. As shown in Figure 5.1, the sterile, hybrid and the restorer lines of sunflower show two main *atpA* homologous transcripts, ca. 3kb and 1.9kb in size. The larger 3kb transcript is more abundant than the smaller 1.9kb transcript in these cytoplasms. In the fertile line, a 1.9kb transcript is abundant and a low level of a 3kb transcript is apparent. The sizes of these transcripts were calculated by their migration relative to the mt rRNAs and *E.coli* rRNAs used as markers during gel electrophoresis. An identical hybridisation pattern was found using the 3.8kb *EcoRI* fragment which contains the *atpA* gene from sterile sunflower.

Minor RNA transcripts homologous to the 2.2kb *EcoRI* fragment are also present in the sterile and fertile cytoplasms. In Figure 5.1, size markers denote the position of a 1.7kb transcript, present in the sterile, hybrid and restorer lines and a 1.3kb transcript present in the fertile line. Thus, in sterile sunflower, the presence of the 3kb and 1.7kb transcripts is correlated with the sequence divergence 3' to the *atpA* gene found in this cytoplasm. If the *atpA* transcript in fertile sunflower contains a transcribed 5' noncoding region of about 200 bp (by comparison with the *atpA* transcript in *Oenothera* described by Schuster et al. 1986) the 1.9kb transcript in fertile sunflower would terminate approximately at the point described as the probable termination signal (section 4.5).

The low level of the 3kb transcript in the fertile line did not agree with the finding that the 3.8kb *EcoRI* fragment could not be detected at low levels in this cytoplasm. Initially, the faint 3kb hybridisation signal seen in the fertile sample was thought to be due to contamination from neighbouring samples, however subsequent hybridisation experiments showed that an RNA transcript of this size in the fertile cytoplasm gave a reproducible hybridisation with probes covering the 3' coding and flanking region of the *atpA* gene (i.e. the 2.2kb and 3.8kb *EcoRI* fragments, also the maize probe BLSCI). In contrast, deliberately overloading the same sample on a blot probed with a specific 5' coding region maize *atpA* probe (ALXR18) did not result in hybridisation to this transcript in the fertile line (see Figure 5.2). This probe also did not hybridise to the 1.7kb and 1.3kb transcripts indicating they are homologous specifically to the 3' flanking region of the sunflower *atpA* genes and do not represent processed forms of the 3kb and 1.9kb transcripts respectively.
Figure 5.1
Detection of *atpA* Homologous Transcripts in Sunflower Mitochondrial RNA

10μg of total mtRNA from sterile (S), fertile (F), hybrid (H) or restorer (R) type sunflower were fractionated by electrophoresis in a 1.3% w/v agarose gel containing 0.6M formaldehyde. The gel was blotted onto Hybond-N and hybridised with the 32P labelled 2.2kb *EcoRI* fragment from fertile sunflower containing the main *atpA* homology in this genome. Hybridisation, as described in methods, was overnight at 42°C in 5xSSC, 50% v/v formamide, 5x Denhardt’s, 0.1% w/v SDS and 400μg/ml sonicated herring sperm DNA. Autoradiography of the blot revealed labelled major transcripts of 1.9kb (F) and 3.0kb (S, H and R) and minor transcripts of 1.3kb (F) and 1.7kb (S, H and R) as marked. Transcript sizes were calculated by comparison to *E.coli* rRNA markers (not shown).

Figure 5.2 Transcripts Identified by a Maize *atpA* 5’ Coding Region Probe

A duplicate filter of the one shown in Figure 5.1, was hybridised to a 32P labelled DNA probe (ALXR18), containing only the 5’ coding region of the maize *atpA* gene. Autoradiography indicated the major 1.9kb and 3kb *atpA* transcripts, hybridise in fertile and sterile, hybrid and restored sunflower as before. The 1.7kb transcript in the sterile line, and the 3kb and 1.3kb transcripts in the fertile line do not hybridise, showing they do not contain homology to the *atpA* 5’ coding sequence.

Figure 5.3

A 2.6kb Transcript Homologous to the Clone pA212 in Sunflower

The clone pA212 of the 1.2kb *EcoRI* fragment containing the downstream inversion endpoint in sterile sunflower (see Figure 3.23) was labelled with 32P and used to probe a Hybond-N filter containing 10μg mtRNA from sterile (S), fertile (F), hybrid (H) and restorer (R) sunflower fractionated by electrophoresis on 1.3% w/v agarose. Autoradiography revealed that, in addition to the 3kb and 1.9kb *atpA* homologous transcripts, a 2.6kb transcript is present in all the cytoplasms (arrowhead). A slightly reduced hybridisation temperature was used for this probe (37°C) to obtain hybridisation with the *atpA* transcripts and consequently a relatively high level of non-specific hybridisation also occurred.

Figure 5.4

The 2.6kb Transcript is Homologous to the Region 3’ to the Inversion

10μg mtRNA from sterile (S) and fertile (F) sunflower were fractionated by electrophoresis and transferred to Hybond-N as above. This filter was hybridised with the 291bp *AluI-EcoRI* subfragment of pA212 which had been labelled with 32P. This subfragment is entirely within the single-copy region of pA212 (see Figure 4.7) and demonstrates that the 2.6kb transcript does not contain homology to the *atpA* repeat.

151
One explanation of these results is that the hybridisation signal corresponding to the 3kb transcript in the sterile line may actually consist of a doublet of two transcripts. One of these 3kb transcripts encodes the complete \textit{atpA} mRNA and is specific to the sterile cytoplasm. The other 3kb transcript is homologous to only \textit{atpA} 3’ flanking sequence, and is transcribed in fertile sunflower at low level. The presence of the major 3kb \textit{atpA} gene transcript obscures the level of the other 3kb transcript in the mtRNA from the sterile cytoplasm. The minor 3kb transcript in fertile and sterile sunflower could originate from the presence of substoichiometric levels of sequences, homologous to the 3’ flanking region of \textit{atpA}, observed in the mitochondrial genomes of both cytoplasms (see Figure 3.25).

As stated, the same pattern of hybridisation shown in Figure 5.1 was found when the 3.8kb \textit{EcoRI} fragment containing the \textit{atpA} gene from sterile sunflower was used as a probe against mtRNA (data not shown). This indicates that the variant 1.3kb and 1.7kb transcripts, found in sterile and fertile sunflower respectively, are transcribed from sequences which are, at least in part, homologous in these two restriction fragments. Several alternative explanations can account for this observation:

1. The 1.7kb and 1.3kb transcripts may correspond to the orfb sequence in fertile sunflower and the extended version, orfc in sterile which are identical at their 5’ ends (see Figure 4.5a). The low level of orfb which would be present in the sterile genome, as part of the 2.2kb \textit{EcoRI} fragment which occurs at low level, may not be transcribed or may accumulate to only a very low level and not be detected by the mtRNA hybridisations;

2. The 1.7kb transcript may originate from the large ORF (orfd) specific to this region of the sterile genome. The predicted direction of transcription of orfd is opposite to that of \textit{atpA} ORF (orfA, see Figure 4.5a), however, the double stranded probe (i.e. the 2.2kb \textit{EcoRI} fragment) used to investigate transcription from this region, would identify transcripts originating from either strand of the mtDNA. The 1.7kb transcript is sufficiently large to encode the length of orfd which has been sequenced (corresponding to approximately 400 bases) plus an as yet undefined amino terminal region. A long 3’ noncoding region on this transcript could mean it extends beyond the sequence specific to the 3.8kb \textit{EcoRI} fragment, and thus it would contain some homology to the 2.2kb \textit{EcoRI} fragment as found;

3. The presence of the 1.3kb transcript correlates with the orf70 sequence which is found only downstream of the \textit{atpA} gene in the 2.2kb \textit{EcoRI} fragment from fertile sunflower (see Figure 4.5a). Although the same sequence is present in sterile sunflower as part of the repeat within the 1.2kb \textit{EcoRI} fragment, an alternate 5’ flank conferred by the genomic inversion may prevent it being transcribed in this line.
Using the 1.2kb EcoRI fragment from the sterile genome as a probe against mtRNA from the various cytoplasms gave the result shown in Figure 5.3. As expected the atpA homologous transcripts hybridise and, in addition, a 2.6kb transcript common to all cytoplasms is observed. Restriction endonuclease digestion of the 1.2kb EcoRI fragment was used to generate smaller fragments for use as probes to localise the sequence responsible for hybridisation to the 2.6kb transcript. The enzyme AluI cuts the 1.2kb EcoRI at three sites and was used to isolate a 742bp AluI-EcoRI fragment containing only atpA 3' flanking region homology and an EcoRI-AluI of 291bp containing only unique flanking sequence (see Figure 4.7a; the two small internal AluI fragments were not isolated). The 742bp AluI-EcoRI fragment hybridised to the same transcripts as the 2.2kb EcoRI fragment. The 291bp probe covering the single-copy region of the 1.2kb EcoRI fragment hybridises only to the 2.6kb transcript (Figure 5.4). This result is consistent with the finding of an ORF in the region covered by the 291bp probe (orf73) during sequence analysis (Figure 4.7a). However, the size of the ORF described for orf73 represents less than 10% of the total length of the 2.6kb transcript. Thus, orf73 may be part of a longer, transcribed 2.6kb sequence initiated from the adjacent region of the genome (i.e. within the the 3.4kb EcoRI fragment in Figure 3.23). The presence of this 2.6kb transcript in mtRNA from fertile sunflower suggests that the orf73 sequence is probably present within the 0.8kb EcoRI fragment from this genome, and which corresponds, in part, to the 1.2kb EcoRI fragment from sterile sunflower.

5.3 CobTranscription is Unaltered Between Fertile and Sterile Sunflower

A cob coding region probe from maize (M13 640) was hybridised with mtRNA from fertile, sterile, hybrid and restored sunflower and gave the result shown in Figure 5.5. A single 2.4kb transcript was found to hybridise in all cases, indicating a pattern of transcription which is simple and unaltered between the cytoplasms. The cob ORF in sunflower has been reported to be 1194bp long, corresponding to a protein of 397 amino acids (Gallerani et al. 1988). The 2.4kb transcript is therefore more than twice the length required to encode the polypeptide and must include 5' and/or 3' non-translated regions, which comprise approximately 1.2kb of the total length of the mRNA. As shown earlier, the inversion in the sterile sunflower genome includes the region encoding the cob gene. This predicts a switch in the direction of cob transcription, relative to the atpA gene, but the rearrangement does not alter the immediate upstream region of the cob coding sequence. This result indicates that the rearrangement of the genome containing the cob sequence has no effect on the size or abundance of the transcript of the gene in sterile sunflower (the variation in level of hybridisation in the restorer line is due to a loading difference).
5.4 Transcripts for All Other Genes Examined are Identical in Fertile and Sterile Sunflower Mitochondria.

Figures 5.6 to 5.11 show the results of hybridisation of $^{32}$P-labelled probes of several other mitochondrial genes to mtRNA from fertile and sterile sunflower. In most cases single transcripts hybridise to the gene probes used, indicating that in sunflower mitochondrial transcription appears to be simple. Single transcripts, which were identical between the fertile and sterile cytoplasm of sunflower, were found for the coxl (2.2kb), nad5 (2.8kb) and atp6 (1.4kb) genes as shown in Figures 5.6, 5.7 and 5.8 respectively. The coxl and nad5 transcripts are also identical in hybrid and restorer sunflower (the transcript homologies for all other genes were only investigated in mtRNA from fertile and sterile sunflower). The result of the nad5 hybridisation shown in Figure 5.7 indicates a simple transcript pattern for this gene in sunflower, in contrast to the complicated nad5 transcription patterns found in Oenothera and wheat (Wissinger et al. 1988, Bonen et al. 1988).

Earlier mtDNA hybridisation experiments led to the conclusion that sequences homologous to the atp9 gene are present in several locations in the sunflower mitochondrial genome (section 3.6) A major transcript of 800bp hybridised with this probe in mtRNA from both fertile and sterile sunflower (Figure 5.9). Three minor transcripts are also present (600bp, 450bp and 400bp marked by the arrowheads) and even the smallest of these would be large enough to encode the entire atp9 proteins which have been sequenced in other species. The minor transcripts may therefore be processed forms of the larger, more abundant primary transcript. Alternatively, they may be transcripts of an incomplete atp9 gene present in the genome as has been reported for Petunia (Rothenberg and Hanson 1987, Bland et al. 1986).

Despite several attempts the quality of mtRNA hybridisation with a coxl intron specific probe could not be improved. The reported size of the intron in sunflower is 1.3kb (Gallerani et al. 1988) and a transcript of approximately this size (1.1kb) was detected. In addition, a low level of hybridisation to transcripts of 2.4kb and 700bp was observed (arrows in Figure 5.10). The coxl exon probe hybridised to a single transcript of 1.5kb (also shown in Figure 5.10) which is most likely to represent the mature coxl mRNA. The combined sizes of this processed transcript and the largest intron transcript (2.6kb) agree reasonably well with the size of the predicted primary transcript (2.4kb).
Figure 5.5 (page 158)
The cob Transcript in Sterile Sunflower is Similar in Size and Abundance in Fertile, Sterile, Hybrid and Restorer sunflower.
10μg of mtRNA from sterile (S), fertile (F), hybrid (H) or restorer (R) sunflower was electrophoresed through a 1.3% w/v denaturing agarose gel and blotted onto Hybond-N. A probe containing the coding region of the maize cob gene (M13 640) was labelled with $^{32}$P and hybridised to the blot at 42°C, as described in methods. Autoradiography revealed a single 2.4kb transcript in all cytoplasms.

Figure 5.6
The coxl Transcription Pattern in Sunflower Mitochondria
A duplicate mtRNA filter to the one shown in Figure 5.5 was hybridised, under the same conditions, with a probe containing the coxl gene from maize. A single 2.2kb transcript was observed in all cytoplasms.

Figure 5.7
A Single, Major Transcript Hybridises with a nad5 Gene Probe in Sunflower
The filter shown in Figure 5.1 was stripped of the 2.2kb EcoRI fragment probe and rehybridised with a probe containing the nad5 gene from Oenothera. A major 2.8kb transcript hybridises in sterile (S), fertile (F), hybrid (H) and restorer (R). As described in the text, a small part of the probe used contains the rrn5 gene and thus explains the hybridisation to the 5S rRNA in the lower region of the blot.

Figure 5.8
The atp6 Transcription Pattern in Sunflower Mitochondria
10μg of mtRNA from sterile (S) and fertile (F) sunflower was electrophoresed through a 1.3% w/v denaturing agarose gel and transferred to Hybond-N. Hybridisation with a $^{32}$P-labelled probe containing the atp6 gene from maize revealed the presence of a single, transcript of 1.4kb in both cytoplasms (the concentration of the mtRNA from fertile sunflower appears to be less than the intended 10μg but longer exposure of the same filter gave the same hybridisation pattern).
Figure 5.9 (page 158)

**A Single Major Transcript and Several Minor Transcripts are Homologous to the** *atp9 Gene in Sunflower*

10μg of mtRNA from sterile (S) and fertile (F) sunflower was electrophoresed through a 1.3% w/v denaturing agarose gel and transferred to Hybond-N. Hybridisation with a $^{32}$P-labelled probe containing the *atp9* gene from maize gave equivalent patterns of hybridisation in the two cytoplasms. The most abundant transcript is approximately 800bp but up to three smaller, less abundant transcripts are also apparent (marked by the arrowheads they are 600bp, 450bp and 400bp).

Figure 5.10

**Transcripts Corresponding to the Intron and Exon Regions of the** *coxII Gene in Sunflower*

10μg of mtRNA from sterile (S) or fertile (F) sunflower was electrophoresed through a 1.3% w/v denaturing agarose gel and hybridised with maize *coxII* gene sequences corresponding to the intron (Int) or exon 1 (Ex) regions. A low level of hybridisation was found with the intron probe but indications of signals at 2.4kb, 1.1kb and 0.7kb are marked. The exon probe hybridised with a single transcript of 1.5kb.

5.11

**The** *nad1* **Homologous Region of mtDNA in Sunflower is Not Abundantly Transcribed in Etiolated Seedlings**

10μg of mtRNA from sterile (S) and fertile (F) sunflower was electrophoresed through a 1.3% w/v denaturing agarose gel and transferred to Hybond-N. Hybridisation with a $^{32}$P-labelled probe of the *nad1* homologous region from *Oenothera* mtDNA and subsequent autoradiography revealed only a faint transcript signal of approximately 900bp in both cytoplasms.
Only a weak hybridisation signal was obtained when a probe of the *nad1* gene from *Oenothera* was hybridised to sunflower mtRNA (Figure 5.11). This contrasts with the strong hybridisation obtained when the same probe was hybridised with sunflower mtDNA (see Figure 3.13), suggesting that, although this gene is present in the genome, it is not highly expressed in etiolated seedlings. A similar lack of expression of the *nad1* gene has been reported for mtRNA purified from wheat seedlings (Bonen *et al.* 1988).

Table 5.1 summarises the sizes of the transcripts detected in mtRNA from etiolated fertile and sterile sunflower. The indication is that transcription of the *atpA* gene is the only one for which there is evidence of variation between fertile and sterile cytoplasms.

### Table 5.1

**RNA Transcript sizes for several mitochondrial genes in sunflower**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fertile Transcript Size</th>
<th>Sterile Transcript Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>atpA 5' *</td>
<td>1.9</td>
<td>3.0</td>
</tr>
<tr>
<td>atpA 3' †</td>
<td>1.9, 1.3</td>
<td>3.0, 1.7</td>
</tr>
<tr>
<td>atp6</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>atp9*</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>cob</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>coxI</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>coxII (intron)</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>coxII (exon)</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>nad1</td>
<td>0.9?</td>
<td>0.9?</td>
</tr>
<tr>
<td>nad5</td>
<td>2.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>

The sizes of transcripts corresponding to several mitochondrial genes detected in fertile and sterile sunflower are indicated. These sizes were calculated by comparison to *E.coli* ribosomal RNA markers included in 1.3% w/v 0.6M formaldehyde agarose gels used to fractionate mtRNA by electrophoresis. * the 5' *atpA* transcripts noted are those which hybridise with the maize probe ALXR18, specific for the 5' coding region of the gene. † the 3' *atpA* transcripts noted are those which hybridise with the 2.2kb *Eco*RI and 3.8kb *Eco*RI *atpA* containing fragments from fertile and sterile sunflower respectively. • less abundant *atp9* homologous transcripts are 0.6kb, 0.45kb and 0.4kb in both cytoplasms.
5.5 Transcription During Flower Development in Fertile and Sterile Sunflower

In view of the differences in the transcripts associated with rearranged *atpA* region in sterile etiolated sunflower seedlings, a preliminary investigation of transcription of the *atpA* locus during flower development was performed. This was in the hope that it would provide further evidence of some causal link between rearrangement in the mitochondrial genome of sterile sunflower, and the CMS phenotype. An initial aim was to investigate *atpA* transcription specifically in anther tissue. However, the 'simple' approach involving mtRNA hybridisation analysis was precluded by the fact that it was impossible to extract sufficient mtRNA from sterile anthers at all stages of development. Instead, total cellular RNA extracts from flowers were used for semi-quantitative hybridisation analyses with various gene probes.

The large compound inflorescence of the reproductive stage of sunflower development consists of a receptacle to which individual disk flowers or florets are attached. The florets, up to 2000 per inflorescence, are arranged in whorls, which mature centripetally at a rate of 1-3 whorls per day. The flowering period of an individual floret lasts only 2-4 days. Due to the protogynous flowering character which sunflowers show, the anthers elongate before the stigma and usually recess, within the first day, before the stigma is fully extended. The stages of flower development used for the RNA extractions are shown in Figure 5.12. Developmental categories, defined on the basis of the length of individual floral buds, were classed as 'young' (less than 2mm measuring from the point of attachment to the receptacle to the petal tips), 'mid' (2-5mm) and 'old' (5-8mm).

Light and electron microscopic examinations of the floral buds used for RNA extraction indicated that the 'young' stage of the fertile and sterile tissues appeared similar in their ultrastructure. In both cases, the stage of development shown by the 'young' buds was found to coincide with prophase of pollen mother cell meiosis within the anthers (Professor Hugh Dickinson, personal communication). These microscopic examinations also indicated that the tapetum is present as a unicellular cell layer investing the meiocytes, in both fertile and sterile anthers from this stage. Thus, only preliminary differentiation of the tapetal layer has occurred in the floral buds by the time they reach the 'young' stage used for transcript analysis increasing the likelihood that the selected stages encompass the point when abnormal reproductive development begins in sterile sunflower.

160
Sunflower plants, grown to maturity in controlled environment growth rooms, were used for harvesting of floral buds from inflorescences at various stages of development. Three main stages of floral bud development, from both fertile and sterile sunflowers, were used and classified as:

(i) 'Young' - floral buds up to 2mm in length (from receptacle attachment site to petal tips);
(ii) 'Mid' - floral buds between 2 and 5mm in length;
(iii) 'Old' - floral buds 5-8mm in length.

At the oldest stage, stigmas were elongated and in fertile flowers pollen production was apparent. The length of the stigmas were not included in the measurements of bud length as this would have introduced deviation in the stages compared between the two phenotypes. The photographs opposite show representatives of the various flower developmental stages used- (a) 'young' stage (b) 'mid' stage from fertile or sterile sunflower (there is little externally visible difference between the phenotypes until the latter part of the mid stage) (c) 'old' stage from fertile sunflower and (d) 'old' stage from sterile sunflower.
5.6 Quantitative Variation in Total RNA and mtRNA in Fertile and Sterile Anthers

In order to gain some measure of the quantitative variation in the mitochondrial RNA transcripts in flowers from fertile and sterile sunflower, total RNA was extracted from 1g of anther tissue from fertile and sterile buds at the 'mid' stage of development. The entire RNA extraction from the anther types was split into two aliquots, fractionated by electrophoresis and blotted onto Hybond-N to create duplicate filters. The filters were then hybridised with end-labelled mtDNA purified from either fertile or sterile etiolated seedlings (Figure 5.13). The level of hybridisation was much lower in sterile sunflower reflecting the lower amount of total RNA present. Large amounts of pollen were not present in the fertile anthers at this stage and so do not contribute greatly to the total RNA extracted. The cytosolic and chloroplast rRNAs hybridise with the total mtDNA probe and similar cross-homologies may be partly responsible for hybridisation to some of the other transcripts which are labelled. By scanning the autoradiograph shown in Figure 5.13 in a densitometer to quantitate the relative hybridisation signals, the level of the cytosolic ribosomal RNAs in tissue from sterile sunflower was estimated to be 50-fold lower than from fertile. This quantitative variation makes it difficult to assess the extent of qualitative differences in the pattern seen with the mtDNA probes from sterile or fertile seedlings. However, some lower molecular weight signals do appear to show a comparable abundance in the two lines (arrows in Figure 5.13).

5.7 Mitochondrial Transcripts are Most Abundant During Early Flower Development

Specific gene probes were used to examine whether mitochondrial transcription varied detectably with development in fertile flowers. Probes for atpA and cob were hybridised to RNA from flower tissue at the young, mid and old stages of development. Figure 5.14 shows the results of probing equivalent amounts (10μg) of total RNA from hybrid sunflower with probes corresponding to the 5' coding regions of the atpA and cob genes. The steady state levels of transcripts corresponding to both atpA (a) and cob (b) are greatest in the young tissue and shows a progressive decrease with age. Densitometer scans of the autoradiographs indicate that the abundance of both the 3kb and 1.9kb atpA transcripts present in this line, falls some 2.5 fold between the young and mid stages and a further two fold between mid and old stages. For the 2.4kb cob transcript, the estimated reductions were 1.3 fold between young and mid stages, and 1.4 fold between the mid and old stages. Thus, with increasing maturity of the tissue, the steady state levels of the both the atpA and cob transcripts decreases.
Figure 5.13 Detection of transcripts homologous to mtDNA in total anther RNA from fertile and sterile sunflowers

Total RNA from 1g of anther tissue from fertile (F) or sterile (S) sunflowers was electrophoresed through 1.3% w/v agarose and blotted onto nylon membrane. Duplicate filters were hybridised with mtDNA from etiolated fertile (a) or sterile (b) seedlings end-labelled with $\gamma^{32}$ P. Arrows indicate transcripts of comparable abundance in the two lines.

Figure 5.14 Variation in the levels of the atpA and cob transcripts occur during flower development.

10µg of total RNA from florets of hybrid sunflower at the young (Y), mid (M) or old (O) stages of flower development (described in the main text) were electrophoresed as above and hybridised with probes for atpA (a) or cob (b)
As shown in Figure 5.4, the 1.2kb EcoRI fragment (clone pA212) from sterile sunflower hybridises not only to the *atpA* transcripts but also to a 2.6kb transcript which is present in both fertile and sterile lines. This fragment was used as a probe against total RNA from flowers at the mid stages of development in the fertile (842B) and sterile (842A) lines and all three stages of development of *Helianthus petiolaris*, the fertile line which is the progenitor cytoplasm of the sterile line (shown in (a), Figure 5.15). The photograph of the stained gel in Figure 5.15b shows equivalent amounts (10μg) of total RNA were present in each track of the gel used to produce the filter shown in (c). A very high level of the 2.6kb transcript is present in the RNA from young *H.petiolaris* flowers. The level of the 2.6kb transcript is decreased in RNA from the mid stage and old stages (5.8 fold reduction was estimated for both, relative to the young stage). The mid stages of the sterile and fertile flowers show levels of the 2.6kb transcript between those of the young and mid stage of *H.petiolaris* flowers.

The result shown in Figure 5.15c also demonstrates that the *H.petiolaris* cytoplasm like the 'normal' fertile (which contains the *H.annuus* cytoplasm) does not contain the 3kb *atpA* transcript found in the sterile line (which contains the *petiolaris* cytoplasm). Thus expression of the 3kb *atpA* homologous transcript in the sterile line occurs only in the presence of the *H.annuus* nucleus. This supports the proposal that the *atpA*-linked rearrangement, and consequent altered transcription, are the consequence of induced genomic recombination of the *petiolaris* mtDNA under the influence of a foreign, *annuus* nucleus. However it remains to be confirmed, by mtDNA hybridisation analysis, that *H.petiolaris* contains a mitochondrial genome containing predominantly the 2.2kb EcoRI type *atpA* locus characteristic of the *H.annuus* fertile line.

An *atp9* gene probe was included in the same RNA hybridisation experiment shown in Figure 5.15 (arrow). This gene was used for comparison because the transcript had previously been shown to be small (Figure 5.9) and so would not confuse interpretation of hybridisation to the 1.2kb EcoRI fragment probe. The level of the *atp9* transcript decreased 2.3-fold between young and old stages of *H.petiolaris* flowers, in a trend comparable with the *cob* transcript noted earlier. The low level of hybridisation with the *atp9* probe may reflect a low steady state level of this transcript in the tissues examined. The *atp9* and 1.2kb EcoRI fragment probes used in the mixed hybridisation experiment were of similar specific activity (10⁹ dpm per μg starting DNA). However the heterologous nature of the *atp9* gene (maize mtDNA) may have reduced the level of hybridisation relative to the homologous 1.2kb EcoRI fragment probe from the sunflower genome under the hybridisation conditions used.
Figure 5.15
The 2.6kb Mitochondrial Transcript Homologous to Clone pA212 Decreases during floral maturation

Total RNA from three stages of flower development of Helianthus petiolaris, (a) the fertile progenitor of the sterile line was prepared as described in methods. 10µg of RNA from the three stages: young (Py), mid (Pm) and old (Po) plus the mid stages from sterile (S) and (F) lines was fractionated by electrophoresis through an 1.3% w/v denaturing agarose. The stained gel is shown in (b) and shows the 5µg of total RNA from E.coli (E) used as ribosomal markers. The gel was blotted onto Hybond-N and hybridised with 32P-labelled probes of the sunflower 1.2kb EcoRI fragment, contained in clone pA212, and the coding region of the maize atp9 gene. All lines show the 800bp transcript homologous to the atp9 gene (arrow). The transcripts which hybridise in sterile sunflower are 3kb, 2.6kb and 1.9kb transcripts (as shown in Figure 5.3) and the transcripts which hybridise in fertile and H.petiolaris are the 2.6kb and the 1.9kb. (The faint high molecular weight band in the Py track corresponds to nonspecific hybridisation to the large rRNA which occurred with this sample and which was removed by a stringent wash of 0.1x SSC at 42°C for 15 minutes). A very high steady state level of the 2.6kb transcript is present in young tissue from H.petiolaris, but which decreases in the later developmental stages.
Figure 5.15a  *Helianthus petiolaris* - the CMS progenitor
As it appeared to be highly expressed, and therefore easily detectable, the 2.6kb transcript was used to determine more precisely when the major change in the level of mitochondrial transcription occurs in flower tissue from hybrid sunflower. At the time of the experiment, hybrid sunflower plants were available which contained florets at the required stages of development and so were used for the total RNA extraction. The young and mid stages, described above, were further separated into two categories to produce a total of four categories of harvested flowers: (i) up to 1mm in length (very young, VY), (ii) 1-2mm (young, Y) (iii) 2-3mm (young-mid, YM) (iv) 3-5mm (mid M). Figure 5.16 shows a stained gel containing 10µg of total RNA from these four stages (a) hybridised with a probe specific for the 2.6kb transcript (the 291bp AluI-EcoRI described in Figure 4.7). The short exposure of the autoradiograph (i) indicates that the level of the 2.6kb fragment decreases 5-fold (estimated by densitometric scanning) between the Vy and Y stages followed by further reductions of ca. 0.4 fold for both of the two later stages.

5.8 Analysis of Mitochondrial DNA during Flower Development

Total cellular DNA was also purified from the same four stages of development used in the RNA preparations from hybrid sunflower, described above. 10µg of total DNA from each of the four stages was digested to completion with EcoRI, fractionated by gel electrophoresis and transferred to Hybond-N. Figure 5.17 shows an autoradiograph of the filter which has been probed with the 1.2kb EcoRI fragment which hybridises to the 2.6kb transcript. The mtDNA fragments which hybridise to this probe are as expected, the 1.2kb EcoRI itself, the 3.8kb EcoRI atpA containing fragment and the 0.8kb EcoRI fragment (all as previously shown, see Figure 3.24). These are present at approximately equal levels at all four stages of flower development (intensity differences between very young and young stages reflect slight loading variations). Thus, the decrease in the steady state level of this mitochondrially encoded transcript is a not a reflection of the copy number of the DNA sequence which encodes it. The variation in the steady state level of the 2.6kb transcript at various stages of flower development may be due to a differential rate of transcription or a higher rate of transcript turnover and/or degradation with increasing maturity of the flowers.

The hybridisation with genomic DNA shown in Figure 5.17 suggests that the level of mtDNA, relative to total DNA, does not alter during the stages of flower development which were investigated. The reduction in the steady state levels of the atpA, cob, atp9 and 2.6kb transcripts are therefore not the consequence of a reduction in the amount of mtDNA per given weight of tissue. There is some suggestion that different gene transcripts may accumulate to different levels in young flower tissue. This may be a result of differential expression leading up to and during normal flower development in sunflower.
Figure 5.16
Decrease in the 2.6kb Transcript Abundance Begins at an Early Stages of Flower Development

Total RNA was extracted from flowers of hybrid sunflower which were less than 1mm (Vy), 1-2mm (Y), 2-3mm (Ym) or 3-5mm (M) in length. 10µg of total RNA from each stage was electrophoresed through a 1.3% w/v denaturing agarose gel and visualised by ethidium bromide staining (a). The gel was then blotted to Hybond-N and the filter hybridised with a 32P-labelled probe of the 291bp AluI-EcoRI mtDNA fragment from sterile sunflower previously shown to hybridise to the 2.6kb transcript. Two different autoradiographic exposures are shown corresponding to (i) 8 hours and (ii) 3 days.

5.17
The Level of Mitochondrial DNA Does Not Alter During the Stages of Flower Development Investigated

The same four stages of flower development used for total RNA extraction shown in Figure 5.16, were also used for preparation of total cellular DNA. Approximately 10µg of DNA from each stage (Vy, Y, Ym, M as described above) was digested with EcoRI and the products of digestion separated by electrophoresis in a 0.7% w/v agarose gel. An EcoRI digest of the 3.8kb EcoRI atpA gene clone from sterile sunflower (pA38) was also included in the gel (lane marked 3.8) as a positive control for subsequent hybridisation. The DNA was transferred Hybond-N and hybridised with the 32P-labelled 1.2kb EcoRI mtDNA fragment from sterile sunflower (clone pA212) and autoradiographed. The extent of hybridisation is equivalent for each sample (the slight variations reflect loading differences precisely).
5.9 A Comparison of Nuclear and Mitochondrial Transcription in Fertile and Sterile Sunflower During Flower Development.

In order to look for evidence of variation in mitochondrial genome expression between the fertile and sterile flower development, changes in the amount of the 2.6kb transcript were used as a 'marker' of expression of the mitochondrial genome. The aim was to determine how the level of this transcript compared with cytosolic rRNAs, used as a marker for nuclear genome expression. Total extractable RNA from 1g of flower tissue from the young, mid and old stages of fertile or sterile sunflower was fractionated by electrophoresis through denaturing agarose as shown by the stained gel in Figure 5.18a. By the latest stage of development, flowers from fertile sunflower contain significant amounts of pollen which contain RNA that will contribute to the total extractable RNA from this tissue. In order to make a feasible comparison with the sterile tissue, attempts were made to remove most of this pollen just prior to harvesting the fertile tissue, by dusting it from the inflorescence heads. In addition, ice-cold grinding buffer was used to wash the freshly harvested tissue several times to remove much of the remaining pollen. Obviously, not all traces of pollen could be removed and it will make some contribution to the RNA levels found at the later stage in the fertile tissue.

The young, sterile flower buds were found to contain more cytosolic rRNA per gram of fresh tissue than fertile flowers of the same length. Beyond this stage, there is a marked reduction in the level of cytosolic rRNA present in sterile sunflower. In contrast, the greatest level of cytosolic rRNA is found in the mid stage fertile tissue, which is followed by a reduction in the older flowers. It appears that the timing of synthesis and/or degradation of cytosolic rRNA in flower tissue of the sterile line may differ to that of the fertile flowers. Variation, from both natural and experimental sources, in the harvested material, cannot be discounted. However, the levels of cytosolic rRNA extracted from the mid stages of the fertile and sterile lines were very similar in several different preparations, using the same bud-length criteria, suggesting that the developmental classification was reproducible.

The gel shown in Figure 5.18a was blotted onto Hybond-N membrane and hybridised with the 291bp AluI-EcoRI fragment probe to obtain an estimate of the changes in the level of the 2.6kb transcript. The result of the hybridisation is shown in Figure 5.18b and indicates a number of interesting points:
Figure 5.18
Cytosolic rRNA Levels and mtRNA Levels are Different in Sterile and Fertile Sunflowers at Similar Stages of Development.
Total extractable RNA from 1g of flower tissue from fertile (F) or sterile (S) sunflower at the young (Y), mid (M) and old (O) stages (as described in the text) was compared by separation on 1.3% w/v denaturing agarose gel stained with ethidium bromide (a). An indication of the level of mitochondrial RNA levels was provided by transfer of the gel to Hybond-N and hybridisation with a $^{32}$P-labelled probe of the 291bp AluI-EcoRI fragment to detect the 2.6kb transcript. The corresponding autoradiograph is shown in (b). Note that the highest level of the mitochondrial 2.6kb transcript is found in the young fertile flowers, before the increase in cytosolic rRNA occurs. The level of the 2.6kb transcript in young sterile sunflower is less than in the fertile tissue from the corresponding stage of development.
Figure 5.18

(a)

(b)
(i) Changes in the level of the 2.6kb mitochondrial transcript do not follow those in the cytosolic rRNA in either fertile or sterile tissue. Between the young and mid stages in fertile flowers (lanes 4 and 5 respectively), when extractable cytosolic rRNA per gram of tissue increases considerably, the mitochondrial 2.6kb transcript has already begun to decrease.

(ii) In the young sterile flowers the level of the 2.6kb transcript is comparable with the reduced level found in the mid stage fertile flowers.

Thus, in fertile tissue, a high steady state level of a mitochondrial transcript appears to precede an increase in cytosolic rRNA synthesis. The young stages of sterile flowers which were investigated may be beyond the point in development comparable to the young fertile stage analysed. A peak in the level of the 2.6kb mitochondrial transcript may occur in the sterile buds at an earlier stage than was examined, i.e. also preceding the cytosolic rRNA increase in this phenotype. If this is the case then, for the purpose of further molecular investigations, a classification on the basis of bud length will not be suitable as it obviously does not reflect similar stages between the fertile and sterile phenotypes.

An alternative explanation for this result is that the development in the two phenotypes is not comparable, with respect to mitochondrial transcription, beyond the earliest stages of floral bud differentiation. A peak in the steady state level of the 2.6kb transcript, comparable to that in fertile tissue, may not occur at all in the sterile flowers. If this is true of other mitochondrial gene transcripts, then despite a high steady state level of cytosolic rRNA, the youngest sterile sunflowers examined are already showing a reduced level of expression of the mitochondrial genome.

5.10 Conclusions

The results of the mtRNA hybridisation experiments presented in this chapter have provided additional indications of molecular variation between fertile and sterile sunflower with respect to expression of their mitochondrial genomes. The evidence indicates that an extended 3kb \textit{atpA} homologous transcript, produced in sterile sunflower, is due to rearrangement downstream of the gene which has removed the putative signal sequence for termination of the normal 1.9kb \textit{atpA} mRNA. An unaltered \textit{atpA} protein could however, be translated directly from this elongated message, since the coding region is unaltered by the chimeric 3' sequence. A 1.3kb transcript in fertile sunflower and a 1.7kb transcript in sterile sunflower, both homologous to the 3' flanking region of the \textit{atpA} gene, provide a further indication of altered expression of the mitochondrial genome between the two phenotypes. All other mitochondrial genes investigated show transcription patterns which are unaltered between fertile and sterile sunflower.
During an analysis of transcription in flowers the abundance of several mitochondrial transcripts was found to decrease with maturation of the reproductive tissue. The level of mtDNA relative to total genomic DNA is constant throughout the stages of flower development which were examined. This indicates that variation in the mtRNA fraction is due to differential transcription and/or stability rather than changes in copy number of the mitochondrial genome. Results obtained suggest that a high level of mtRNA gene expression precedes a large increase in cytosolic rRNA synthesis in fertile sunflower. Sterile flowers at the youngest stage examined show variation in both cytosolic rRNA and mitochondrial transcript levels in comparison with fertile flowers. The reproductive buds of fertile and sterile sunflowers are of comparable lengths at this stage, with the aberrant nature of the anthers becoming obvious only by the 'mid' stage of development. Thus, the molecular criteria indicate that abnormal development in sterile sunflowers is underway by the time physical breakdown of anthers is apparent. Although the results described do not represent a complete study of mitochondrial gene transcription during microsporogenesis, the indication is that the correct regulation of nuclear and mitochondrial gene expression is lacking within immature florets of sterile sunflower.
CHAPTER SIX
MITOCHONDRIAL PROTEIN SYNTHESIS
IN SUNFLOWER

6.1 Introduction

The incorporation of $[^{35}S]$-methionine into polypeptides synthesised \textit{in organello} can be used to investigate translation in mitochondria isolated from plant tissue. Such mitochondrial 'translation profiles' have been used to detect the synthesis of variant polypeptides by mitochondria isolated from fertile and sterile cytoplasms of maize, sorghum, and a number of other species (Forde and Leaver 1980, Bailey-Serres et al. 1986). In CMS-T maize, the variant 13kDa translation product is encoded by a chimeric open reading frame generated by aberrant intragenomic recombination events (Dewey et al. 1986 and see section 1.9.4). In sorghum, as a result of recombination in the carboxy terminal region of the \textit{coxl} gene, an extended version of the COXI protein is synthesised in mitochondria isolated from the sterile (9E) cytoplasm (Bailey-Serres et al. 1986). On the basis of these observations in other species, experiments were designed to determine suitable conditions for \textit{in organello} protein synthesis by mitochondria isolated from sunflower and to look for evidence of variation between the mitochondrial translation products from fertile and sterile conferring cytoplasms.

6.2 The Problem of Bacterial Contamination

Early attempts to obtain \textit{in organello} protein synthesis with sunflower mitochondria were hampered by a high level of bacterial contamination. This was indicated by a high rate of incorporation of $[^{35}S]$-methionine when an acetate energy source, which mitochondria cannot utilise as an oxidisable substrate, was included in the incubation, and also by the presence of numerous labelled polypeptides as opposed to the more normal 15-20 polypeptides synthesised by plant mitochondria. Special care was therefore required when growing material for mitochondrial incorporation experiments to ensure that everything which came into contact with the seedling tissue was sterilised by autoclaving, including water for planting and rinsing, vermiculite growth medium and foil to cover growth trays. These precautions decreased the level of contamination slightly, but it was apparent that the majority of bacterial contamination actually comes from within the seed, which could only be reduced by surface sterilisation with sodium hypochlorite both before planting and after harvesting of the tissue. The \textit{in organello} incorporations therefore had to be performed with the less than ideal material and care in interpretation of the results was required.
6.3 The Effect of Energy Source on Protein Synthesis by Isolated Sunflower Mitochondria

Mitochondria isolated from dark grown seedlings of fertile (842B), sterile (842A), hybrid (844H) and restorer (843R) sunflower were incubated with $^{35}$S-methionine in the presence of a CP/CPK 'energy generating system' or succinate and ADP (as described in methods). In the presence of the exogenous energy generating system, the normal spectrum of plant mitochondrial translation products was obtained with mitochondria isolated from all the cytoplasms (Figure 6.1). When succinate (an oxidisable substrate) and an ADP energy source was used, reasonable levels of $^{35}$S-methionine incorporation were obtained with the fertile (F), hybrid (H) and restorer (R) cytoplasms. However, a poorer level of incorporation with succinate and ADP was consistently found with mitochondria isolated from the sterile (S) cytoplasm (Figure 6.1).

The observations regarding succinate were made with mitochondria isolated from etiolated seedlings and cannot be extrapolated to other tissues. Preliminary comparisons of protein synthesis by sunflower mitochondria isolated from hypocotyls or cotyledons indicate there is some variation in the polypeptides synthesised in the different tissues (data not shown). The variation, although subtle, is both quantitative and qualitative, as the levels of several polypeptides alter slightly between the tissues and some appear to be specifically synthesised in one tissue type. In addition, the nature of the supplied energy source (ATP alone or CP/CPK and ATP) or substrates (malate, pyruvate or succinate) can subtly influence the mitochondrial translation profile produced and for some proteins appears to enhance the tissue specific differences.

6.4 A Comparison of Polypeptides Synthesised by Mitochondria from Fertile and Sterile Sunflower

Preliminary experiments suggested that mitochondria from sterile sunflower synthesised an increased level of a polypeptide with a molecular weight of ca. 17-19kDa when compared to that synthesised by mitochondria from fertile sunflower. By use of a 12-20% (w/v) SDS-polyacrylamide gradient gel to increase resolution of polypeptides in this region of the gel, this apparent 'increased level' was actually found to be due to the presence of a doublet containing two polypeptides, only one of which was synthesised in the fertile line (Figure 6.1)

Further experiments using the high resolution gel system indicated that the additional polypeptide had an estimated molecular weight of 15kDa (and hereafter will be referred to as 'the variant polypeptide' or 'the 15kDa polypeptide'). This variant protein was synthesised by mitochondria isolated from several other lines, including the line sterile line 841A but not in the
isogenic fertile line 841B (Figure 6.2, lanes S1 and F1 respectively). It was also found to be synthesised by the hybrid (H) and restorer (R) lines of sunflower (Figures 6.1 and 6.2).

These results show that synthesis of the 15kDa polypeptide correlates with the presence of the petiolaris cytoplasm (in 842A and 841A) and the rearrangements of the mitochondrial genome described for the 842A cytoplasm in preceding chapters. However, the expression of the 15kDa polypeptide does not appear to be affected by the presence of the restoring nuclear background (present in 844H and 843R). Synthesis of the α-subunit of the F1 ATP synthase complex, showing a molecular weight of ca. 58kDa and marked ATPA on Figure 6.1, appears to be the same in mitochondria from all the cytoplasms investigated. Synthesis of a polypeptide of ca. 34kDa, which probably represents the translation product of the cob gene, is also identical between cytoplasms (marked COB on Figure 6.1).

Apart from the 15kDa polypeptide, there are other, reproducible but less obvious quantitative, variations in the synthesis of lower molecular weight polypeptides of ca. 9kDa and 13kDa. These differences are particularly noticeable in mitochondria from hybrid and restorer lines when succinate is used as an oxidisable substrate (marked by arrows on Figure 6.2). The hybrid cytoplasm also shows increased synthesis of a protein of approximately 18kDa molecular weight (also marked). Although interpretation of the translational profiles shown in Figure 6.2 is difficult, due both to a generally higher level of incorporation in the hybrid and restorer lines and to several subtle variations between samples, the extent of the quantitative variation appears to be greatest for the 9kDa and 13kDa polypeptides described. The smaller of these shows migration in the SDS-polyacrylamide gel characteristic of the DCCD-proteolipid binding protein or subunit 9 of the F0 ATPase synthase (in Figure 6.1 this ATP9 subunit protein has migrated off the gel).
Figure 6.1
Polypeptide Synthesis by Mitochondria Isolated from Fertile, Sterile, Hybrid and Restorer Sunflower

Mitochondria isolated from 842B fertile (F), 842A sterile (S), 844H hybrid (H) and 843R restorer (R) sunflower were incubated for 90 minutes at 25°C in a medium containing 35S-methionine as described in methods (section 2.2.17). The energy sources used to support in organello protein synthesis were either creatine phosphate /creatine phosphokinase plus ATP (CP/CPK) or 10mM succinate and ADP (succinate). Mitochondrial polypeptides were fractionated by electrophoreses in a 12-20% w/v SDS polyacrylamide gradient gel and labelled proteins were detected by autoradiography. The molecular weight markers are shown in kilodaltons (kDa) and the variant ca.15kDa, present in sterile, hybrid and restorer lines is indicated ('15kDa') The translation products corresponding to the α-subunit of the F1-ATP synthase, and the subunit b of the apocytochrome are indicated as ATPA and COB at ca. 58kDa and 34kDa respectively.
Figure 6.1

![Image of gel electrophoresis]

- ATPA
- COB
- '15kDa'
- CP/CPK
- SUCCINATE
Polypeptides synthesised by mitochondria from two fertile lines, 841B (F1) and 842B (F2) and two sterile lines, 841A (S1) and 842A (S2), plus 844H hybrid (H) and 843R restorer (R) lines, were labelled with $^{35}$S-methionine as described in methods (2.2.17). The energy sources used to support *in organello* protein synthesis were creatine phosphate /creatine phosphokinase plus ATP (CP/CPK, for all lines) or 10mM succinate and ADP (succinate, for F2, S2, H and R). Total mitochondrial protein was fractionated by electrophoresis through a 12-20% w/v SDS polyacrylamide gradient gel and labelled proteins were detected by autoradiography. The variant 15kDa polypeptide which is synthesised in both sterile lines, and the hybrid and restorer lines, is indicated. Among the labelled translation products noted with the oxidisable substrate succinate, increased levels of a ca.9kDa polypeptide, probably subunit 9 of the $F_0$ ATPA synthase complex (marked ATP9) and a ca.13kDa polypeptide (13kDa) are seen in the hybrid and restorer lines. In addition, the hybrid appears to shows increased synthesis of a larger 18kDa (18kDa) protein with this substrate.
Figure 6.2

[Image of a gel with molecular weights in kDa and protein bands marked]

- 18kDa
- 15kDa
- 13kDa
- ATP9

CP/CPK → SUCCINATE
6.5 Conclusions

The results presented in this chapter provide further evidence of variation in the expression of the mitochondrial genomes of fertile and sterile sunflower. It has been found that the level of different mitochondrial proteins synthesised in organello can vary with the cytoplasm, the tissue source and the particular substrate/energy conditions used. A variant 15kDa polypeptide synthesised by mitochondria from sterile, hybrid and restorer sunflower is absent from the translation products synthesised by mitochondria from the fertile cytoplasm.

The synthesis of the 15kDa polypeptide in sterile sunflower correlates with the genomic reorganisation in the vicinity of the \( atpA \) gene and the associated extended 3kb \( atpA \) mRNA, described in the preceding chapters. However, no detectable variation in the in organello synthesis of the \( \alpha \)-subunit of the F\(_1\) ATP synthase occurs in sterile sunflower. This supports the earlier proposal that the chimeric region at the 3' end of the 3kb \( atpA \) transcript, which encodes the protein in sterile sunflower, does not modify translation of this mRNA. Also as expected from the transcriptional data, expression of the \( cob \) gene, at the level of translation, is unaffected by its linkage to the region of the mitochondrial genome which is rearranged in sterile sunflower.

The sequence analysis in described in Chapter Four indicates that an ORF encoded by the region downstream of the \( atpA \) gene of the sterile genome, orfd, is part of the rearrangement proposed to have arisen by an insertion event. Orfd could encode a polypeptide of at least 14.4kDa (the amino acid composition of this sequence is shown in Figure 4.5a). As the sequence obtained for orfd does not include an N-terminal methionine initiation codon, this ORF may be longer and encode a larger polypeptide which could be a candidate for the novel 15kDa polypeptide synthesised in sterile sunflower.

A second ORF in the mitochondrial genome of sterile sunflower (orfc) is a modified, extended version of a sequence present in fertile sunflower (orfb, see Figure 4.5a) and includes a putative ribosome binding site in its 5' flanking region. The molecular weight of the protein sequence predicted from orfc is 10.2kDa (the predicted polypeptide from orfb is only 4.9kDa). The extent of SDS-binding to proteins can vary depending on their amino acid composition and affect the interpretation of results from SDS-polyacrylamide gel electrophoresis used (See and Jackanowski 1989). A lack of negative charge, and/or insufficient denaturation due to reduced SDS binding, can contribute to an anomalously high estimate of the molecular weights of particular proteins. Thus, although smaller, the predicted polypeptide encoded by the orfc sequence could also be the 15kDa polypeptide found in the sterile cytoplasm.

The level of the 15kDa polypeptide in the restored and hybrid cytoplasms (phenotypically fertile) is similar to that in the sterile cytoplasm. However, this does not necessarily preclude a causal association of the 15kDa polypeptide with...
the CMS phenotype. There are indications that variations in mitochondrial protein synthesis occur in the line restored to fertility and these may be associated with suppressing the sterile phenotype. Alternatively, at the molecular level, the mechanism of fertility restoration in sunflower may be complex and involve other, as yet, undefined variables. The nuclear restoring gene must encode a gene product which can in some way 'compensate' for any mitochondrial malfunction caused by the expression of variant polypeptides. This compensation may occur by specific binding with a variant polypeptide, to inhibit any association with other protein complexes, thereby reducing the detrimental affect. If this type of mechanism is involved in fertility restoration in sunflower, it would not be apparent from a comparison of the levels of protein synthesis between mitochondria isolated from restored and non-restored cytoplasms.
CHAPTER SEVEN

DISCUSSION

7.1 Sunflower - An Ideal System for Studying Cytoplasmic Male Sterility

The preceding chapters have presented the observations made from a comparison of the organisation and expression of the mitochondrial genomes of fertile and of sterile sunflower, the aim being to determine the precise molecular variation within mitochondria of sterile sunflower which is responsible for the CMS phenotype. The sunflower offers several advantages for a detailed investigation of the mitochondrial genome mutations which cause cytoplasmic male sterility.

Restriction endonuclease digestion and hybridisation analysis show that the organisation of the mitochondrial genomes of fertile and of sterile sunflower is very similar. For the majority of genes investigated, no difference in the genome organisation in the immediate vicinity of their coding regions was detected. Restriction endonuclease fragments, homologous to eleven different mitochondrial genes, are identical in the fertile and the sterile sunflower lines which were examined. Hybridisation of an $\textit{atpA}$ gene probe to $\text{EcoRI}$ digested mtDNA from fertile and sterile cytoplasms, identified two DNA fragments with altered migration indicative of genome reorganisation (the 2.2kb and 0.8kb $\text{EcoRI}$ fragments in fertile sunflower and the 3.8kb and 1.2kb $\text{EcoRI}$ fragments in sterile sunflower). An additional fragment, which is altered between the lines (the 2.1kb $\text{EcoRI}$ fragment in fertile sunflower and the 2.4kb $\text{EcoRI}$ fragment in sterile sunflower) is part of the same genome reorganisation.

The similarity between the restriction endonuclease digestion patterns of mtDNA from fertile and from sterile sunflower, and the identification of only limited variation in the genome organisation associated with the $\textit{atpA}$ locus, contrasts with findings in maize for which 1-12 differences per 50 fragments between the mtDNAs of fertile and CMS lines have been reported (Pring and Levings 1978). Most of the rearrangements of the mitochondrial genome in maize CMS lines do not involve coding sequences and thus have no discernable effect on gene expression. This suggests that, at most, only a subset of genomic rearrangements can be causally linked to the CMS phenotype in maize.

As discussed earlier, intramolecular and intermolecular recombination of the genome in plant mitochondria appears to be a frequent process. Genome rearrangements involving duplication and relocation of gene regulatory and/or coding sequences may prove 'advantageous' in some instances, leading to selection of the functional sequences involved. However, it is to be expected
that periodically the creation of molecular diversity will also result in aberrant genome rearrangements which are detrimental to normal mitochondrial gene expression, such as those implicated in CMS.

The similarity between the mtDNAs of fertile and sterile sunflower investigated is somewhat unexpected given that the cytoplasms originate from different species (H. annuus and H. petiolaris in fertile and sterile respectively). This strengthens the likelihood that any variation between the mitochondrial genomes of the two phenotypes will indeed be related to the sterility trait. This thesis presents evidence that variation between the mitochondrial genomes of fertile and sterile sunflower is associated with altered transcription in the two cytoplasms. However, a definitive causal link between the altered expression and the CMS phenotype remains to be established.

7.2 Variant Expression of the Mitochondrial Genome of Sterile Sunflower

Intragenomic recombination within the mitochondrial genome of sunflower has resulted in a mtDNA rearrangement in the region encoding the genes for atpA and cob. It is proposed that homologous recombination across a short, indirectly repeated sequence in mtDNA of fertile sunflower results in inversion of a ca.13kb region to generate the form of the genome characteristic of the sterile line. In addition, one end of the inversion, adjacent to the 3' end of the atpA gene in the sterile genome contains a sequence insertion of approximately 2.3kb. The repeats involved in this recombination mark the endpoints of the genomic rearrangement and correspond to four restriction fragments of altered mobility noted in mtDNA restriction endonuclease profiles. These are the 2.2kb and 0.8kb EcoRI 'progenitor' fragments present in the fertile genome, and the 3.8kb and 1.2kb EcoRI fragments, derived by recombination, which occur in the sterile genome. The 2.2kb EcoRI fragment and the 3.8kb EcoRI fragment contain the main atpA gene coding sequences in fertile and sterile sunflower respectively.

The genomic rearrangement found in the mtDNA of sterile sunflower appears to influence the expression of the atpA gene in this cytoplasm. The atpA mRNA transcript in fertile sunflower is 1.9kb in size. In sterile sunflower the rearrangement has resulted in an altered 3' region of the atpA gene from which an abnormally long 3kb atpA mRNA is transcribed. The 3kb atpA mRNA predominates in the sterile cytoplasm although a detectable level of the 1.9kb transcript found in fertile sunflower, is also present. Sequence analysis has shown that the coding region of the predominant atpA gene in sterile sunflower is unaltered by the genomic rearrangement. However, the DNA sequence in the 3' flank of the gene is completely different from that present downstream of atpA gene in fertile sunflower. A sequence which probably acts as a signal for termination of the atpA transcript in fertile sunflower is missing
from the corresponding region of the \textit{atpA} gene in sterile sunflower. This sequence, consisting of a 27bp inverted repeat has the potential to form a large stem-loop structure and shows some features of the mapped transcript termination sites of other plant mitochondrial genes.

In chloroplast genomes, inverted repeats are located 3' to the coding regions of a number of genes and, by secondary structure formation, may function as transcriptional terminators (Deno \textit{et al.} 1984, Kirsch \textit{et al.} 1986). Recent studies indicate that the stem-loops predicted by these inverted repeats actually function as processing sites and/or to protect mRNA from nucleolytic degradation (Stern and Gruissem 1987). By determining differential stability of plastid mRNAs, such 3' inverted repeats appear to play a role in the regulation of chloroplast gene expression. The inverted repeats located 3' to several mitochondrial genes can be folded into various stem-loop structures and have generally been ascribed roles in transcript termination (Schuster \textit{et al.} 1986, Rothenberg and Hanson 1986). A definition of the requirements for termination, processing and stabilisation of plant mitochondrial gene transcripts will await direct analyses of 3' gene sequences in a functional \textit{in vitro} assay system similar to that developed for chloroplasts.

An analysis of the steady state levels of \textit{atpA} homologous transcripts, and \textit{in organello} translation products synthesised by mitochondria from sterile sunflower indicate that, although altered in size, the level of \textit{atpA} mRNA expression is comparable with that of the fertile line. The 1.9kb \textit{atpA} transcript, additionally present in the sterile line may be translated and contribute to the level of F$_1$ ATP synthase $\alpha$-subunit protein which is produced at an apparently normal level. The conclusion from these results is that the \textit{atpA}-linked rearrangement in the mitochondrial genome of sterile sunflower, is responsible for altered transcription of the gene but does not influence the composition or level of the functional gene product which is expressed in this cytoplasm. The CMS phenotype is therefore unlikely to be related to any malfunction of the F$_1$ ATP synthase complex as a consequence of limited $\alpha$-subunit availability. It should be stressed that these conclusions are made on the basis of the data obtained for etiolated seedling material which may not accurately reflect mitochondrial genome expression at other stages of sunflower development, such as pollen development.

A combination of restriction endonuclease mapping and transcript analysis indicates that, although genome reorganisation in sterile sunflower includes the region encoding the \textit{cob} gene, no affect on expression of this coding sequence arises as a consequence. In mtDNA from fertile sunflower, the \textit{cob} gene is located approximately 8kb downstream of the \textit{atpA} gene and is transcribed from the complementary strand. The inversion in sterile sunflower relocates the \textit{cob} gene relative to \textit{atpA} gene so that the two are approximately
6kb apart and transcribed from the same DNA strand. The level of expression of the \textit{cob} gene would be affected by the rearrangement only if it was regulated by long-range interactions involving sequences located several kilobases up or downstream of the coding region. This does not appear to be the case nor does the relocation of \textit{cob} result in any change to the size of the transcript, or the corresponding protein synthesised in mitochondria from sterile sunflower.

The region of the \textit{atpA} locus which shows sequence divergence in sterile sunflower when compared to fertile contains at least two ORFs in addition to the one which encodes \textit{atpA}. A modified form of a 43 codon sequence present in fertile genome (orfB), has been extended to a 90 codon sequence in sterile sunflower (orfC). OrfC potentially encodes a protein of 10.2kDa, which is over twice the size of the predicted translation product of orfB (4.9kDa). A second ORF (orfD) is present in the region of the 2.3kb insertion in mtDNA of sterile sunflower, beyond the sequence encoding orfC. However, orfD does not contain an ATG start codon and so the 5' terminus of this ORF must be initiated beyond the region sequenced. Translation of the orfD sequence predicts a protein of at least 14.4kDa in size.

The DNA sequence in the region 5' to the ORFs b and c in fertile and sterile sunflower respectively is very similar to a sequence which includes the transcriptional promoters for the \textit{coxI} and \textit{coxIII} genes in \textit{Oenothera}, and the \textit{atp6} gene in tobacco (Hiesel et al. 1987, Bland et al. 1987). Thus, unlike the T-urfl3 gene of CMS-T maize where recombination has placed a functional (atp6) promoter \textbf{upstream} of a chimeric region of the CMS-T genome, in sunflower it appears as if a recombination event has occurred \textbf{downstream} of a pre-existing promoter sequence, although whether this promoter is functional in sunflower remains to be shown. The fact that the putative promoter sequence is reiterated several times within mtDNA from both fertile and sterile sunflower suggests it may be a functional domain which is subject to duplication and relocation by mitochondrial genome rearrangements.

A putative ribosome binding site, thought to be necessary for translation is encoded by the region immediately upstream of the initiator methionine common to both orfB, in fertile, and orfC, in sterile, sunflower. The orfC sequence shares its first ten codons with orfB and an additional eight (i.e. the 18 amino-terminal codons in total) with the orfB sequence described in \textit{Oenothera} by Hiesel et al. 1987. Although no identity has been assigned to these three ORFs, the level of homology implies there is some functional significance to the sequences and supports the proposal that orfB and the chimeric orfC sequence are expressed in fertile and sterile sunflower respectively.

In addition to the \textit{atpA} mRNAs, two other transcripts, homologous to the region of the genome which is rearranged sterile sunflower, show differences in fertile and sterile sunflower. A 1.3kb transcript in fertile sunflower, and a 1.7kb
transcript in sterile sunflower, hybridise specifically with probes for the 3' region of the \textit{atpA} gene and do not represent processed forms of the 1.9kb and 3kb \textit{atpA} mRNAs. The 1.3kb and 1.7kb transcripts are large enough to encode the orfb and orfc sequences respectively, and could initiate in the intergenic region between \textit{atpA} and the ribosome binding site immediately upstream of the start of their ATG initiation codons. However, the 1.9kb and 3kb \textit{atpA} mRNAs in fertile and sterile sunflower would be expected to extend beyond the region containing the orfb and orfc sequences. Indeed the stem-loop proposed as the termination signal for the 1.9kb \textit{atpA} transcript occurs just 22 nucleotides downstream of the end of orfb.

Alternative possibilities for the origin of both the 1.3kb transcript in fertile and the 1.7kb transcript in sterile sunflower can be forwarded. The 1.3kb transcript may correspond to the orf70 sequence located further downstream of \textit{atpA} and orfb sequences in fertile sunflower. The low steady state level noted for this transcript may be due to a reduced rate of transcription compared with \textit{atpA} and/or a reduced stability. The 1.7kb transcript also present at low abundance in the sterile line, may be derived from the orfd sequence located downstream of \textit{atpA} and orfc in this cytoplasm. A third possibility is that both the 1.3kb and 1.7kb transcripts are actually transcribed from regions elsewhere in the genomes, but which share homology with the sequences downstream of the \textit{atpA} genes. If the homology is only partial, the level of hybridisation to the 2.2kb and 3.8kb \textit{EcoRI} probes would not necessarily accurately reflect the levels of the 1.3kb and 1.7kb transcripts in the fertile and sterile cytoplastms respectively.

7.3 Rearrangements in the Sunflower Mitochondrial Genome
- Sublimons and Speculation

The genome of fertile sunflower contains a sequence including the 3' coding and flanking region of the \textit{atpA} gene on a short indirect repeat located approximately 13kb downstream of the intact \textit{atpA} gene. This repeat includes twenty-three codons from the carboxy-terminal sequence of the \textit{atpA} gene which could have acted as the site of homologous recombination which led to the genome inversion found in sterile sunflower. Genomic recombination events involving the 3' flanking region of \textit{atpA} gene have been most extensively characterised in maize and, as the observations made in this system are of potential interest given the findings with sunflower, they will be briefly summarised.

In the fertile (N, normal) cytoplasm of maize the \textit{atpA} gene is located entirely within a 12kb direct repeat sequence (Isaac et al. 1985a). The 3' terminus of the \textit{atpA} gene is located a distance of 650 bases from one end of the repeat and consequently two complete copies of the gene can be represented on the master
circle, which differ in their 3' flanking sequences. MtDNA from CMS-C maize contains a single atpA arrangement, identical to one of those found in N mtDNA. In CMS-S and CMS-T genomes however, novel sequences are located 3' to the single atpA genes contained in these cytoplasms. The alternative atpA arrangements have been designated types 1-4 by Small et al. (1987).

Different maize cytoplasms, within the groupings N and S have been found to vary in the predominant type of atpA arrangement which they contain. In those cytoplasms containing only one predominant atpA type, low levels of two of the other types have been detected (Small et al. 1987). These substoichiometric mtDNA arrangements or 'sublimons' appear to represent recombinant molecules, maintained as part of the entire genome complexity, but present at greatly reduced levels relative to the 'master' circle and predominant subgenomic circles. The study of atpA organisation has been extended to male fertile lines of maize containing the RU cytoplasm which is closely related to the N cytoplasms of modern inbred maize but thought to be ancestral to them (Lonsdale et al. 1988). This has led to a model, based on atpA stoichiometries, for possible evolutionary relationships of subgroups within the designations of N-like (fertile) and S-like (sterile) cytoplasms of maize (Leaver et al. 1988).

Within the RU maize cytoplasm, which contains the atpA 'type 2' arrangement, a restriction fragment specific to this cytoplasm has been found to contain a 181bp repeat homologous to the 3' end of 12kb repeat which occurs in N mtDNA (Small et al. 1989). Recombination across this 181bp repeat predicts the formation of two subgenomic molecules in RU mtDNA. In addition, a 241bp repeat of the 5' terminus of the 12kb repeat is also present in RU mtDNA, and recombination across this repeat would produce two different subgenomic circles. Among the products predicted for subsequent recombination between the alternative subgenomic molecules is a new 'master' circle containing a duplication of the 12kb repeat in an organisation characteristic of N mtDNA. The suggestion is that the 12kb repeat of N mtDNA which contains the atpA genes in present day fertile cytoplasms of maize has arisen by a three-stage recombination process involving sublimon type molecules (Small et al. 1989).

A comparison of the 3' flanking regions of the atpA gene from fertile and sterile sunflower indicates that there is no extended homology between these sequences and either the 181bp repeat or 241bp repeats reported for RU maize. Thus the repeats associated with atpA and which are thought to be active in recombination within the mitochondrial genome of maize, are not related to the ca. 200-300bp repeat postulated to have been involved in the generation of rearrangement in the mtDNA of sterile sunflower (hereafter referred to as 'the 300bp repeat').

However, a survey of available mitochondrial sequences has revealed that part of the 300bp repeat in sunflower is very similar to a sequence which is
reiterated within the mitochondrial genomes of tobacco, *Petunia* and *Oenothera* (Bland *et al.* 1987, Folkerts and Hanson 1989, Hiesel *et al.* 1987). Part of the *atp6* repeat of tobacco is homologous to the repeat upstream of *coxI* and *coxIII* in *Oenothera* and it is this region of shared homology which has now been found to occur in the region 3' to the *atpA* sunflower.

The presence of this conserved repeat element at multiple locations within at least four plant species suggests that it may have a role in the creation of genomic diversity by inducing genomic rearrangement. In sunflower, the main homologies to this conserved repeat are the 2.2kb/0.8kb EcoRI fragments and the 3.8kb/1.2kb EcoRI fragments in the fertile and sterile lines respectively. A low level of the 2.2kb and 0.8kb EcoRI fragments, characteristic of the fertile line, are present in the sterile line. The 3.8kb and 1.2kb EcoRI fragments are not, however, detectable in the fertile line, even at low levels. This suggests that no subliminal molecules containing the arrangement which these fragments represent exist in fertile sunflower. Possibly, the creation of the sterile sunflower line involving transfer of the *H.anuus* nucleus into the progenitor *H.petiolaris* cytoplasm by backcrossing induced novel recombination events within the *H.petiolaris* mitochondrial genome. The organisation of the progenitor *H.petiolaris* genome was not determined, but the transcript analysis indicated the *atpA* transcript pattern in this line was similar to that of the *H.anuus* fertile line.

It is unlikely that the inversion and insertion recombination events which have contributed to the *atpA*-linked rearrangement in sterile sunflower occurred simultaneously and it is also possible that more than one insertion has occurred. If this were the case, a mtDNA restriction fragment, intermediate in size between the 2.2kb EcoRI fragment and the 3.8kb EcoRI fragment and homologous to this 3' region would be predicted. An EcoRI fragment of 3kb present as a 'sublimon type signal' in the mtDNA of sterile sunflower fits these criteria.

Leaver *et al.* (1988) have discussed the significance of sublimons in generating mitochondrial genome diversity. When present at a low level, sublimons would not be expected to have direct phenotypic consequences. As sites for mtDNA rearrangements, they could represent a means for 'cryptic' mitochondrial genome evolution, which would become apparent only upon subsequent amplification of particular molecules. The mechanism by which a previously subliminal molecule becomes a new, dominant form of a genome remains unclear. Selection may favour those molecules containing particular regulatory sequences upstream of essential genes or gene duplications which may be advantageous. Alternatively, the pivotal event in the preferential amplification of a previously subliminal molecule may be the inclusion, by recombination, of an efficient origin of replication.
7.4 Short Tandem Repeats and Mitochondrial Genome Rearrangement

One of the most significant findings with regard to the rearrangement described in the sterile sunflower genome was that a sequence from this region is homologous to one implicated in genome reorganisation in *Oenothera* (Hiesel et al. 1987) and also present in tobacco and *Petunia* (Bland et al. 1987, Folkerts and Hanson 1989). Only part (approximately 140bp) of the 300bp repeat in sunflower, is homologous to the larger 657bp repeat in *Oenothera*. However, as discussed above, the similarity lends support to the proposal that a specific recombination process may be responsible for the duplication and relocation of this particular mtDNA sequence within plant mitochondrial genomes. In sterile sunflower, a sequence within this conserved region or another sequence within the 300bp repeat has been the substrate for homologous recombination causing inversion of a section of the genome. The same region has been the site of secondary recombination event(s) involving a 2.3kb sequence insertion although the exact boundaries of the two events remain to be determined.

Short direct or indirect repeats are found at a number of rearranged sites in plant mtDNAs. In maize mtDNA, tandem 4 and 5bp are common in regions derived by recombination or containing sequences translocated from the chloroplast genome (Pring et al. 1988). Within a maize CMS-T revertant, a tandem 5bp repeat (5' TCTCA3') introduces a termination codon in the ORF encoded by the *T-urf13* gene resulting in the synthesis of a truncated 8kDa associated with reversion to fertility (Wise et al. 1987b). Directly repeated sequences have also been found at the 3' points of divergence between the forms of the *cox1* gene present in Milo (fertile) and 9E (sterile) sorghum (Bailey-Serres et al. 1986). At 30 bp upstream of the point of divergence between these two genes, a 10bp sequence (5' GGAGAACTTC 3') is repeated such that the last C represents the final homologous nucleotide between the two genes (a distance of 11 bases upstream of the stop codon of the Milo *cox1* gene). The same 10bp in the 9E *cox1* gene forms part of a longer 26bp repeat as the following 16bp are repeated in the copies both 5' and immediately 3' to the point of sequence divergence in this genome.

Short direct repeats are also present in sunflower mtDNA in the region of the described sequence rearrangement (see Figure 4.5b). Two copies of a 7bp sequence (5'CTATTTT3') separated by two nucleotides are present in mtDNA from fertile sunflower immediately after the point where sequence diverges in sterile sunflower. A very similar sequence, (5'CTTATTTT3') is located one nucleotide upstream of the point of divergence and is the only copy in this region of the sterile genome. As already stated, this point of divergence is within the ORFs orfb and orfc in the fertile and sterile genomes respectively. Revertants to fertility in sunflower were not available for investigation in this project although they can arise during tissue culture of parental sterile lines. It
would be of interest to examine whether variation in any of these tandem repeats is present in mtDNA from revertants. If any change to the orfc sequence were found within plants which have reverted to fertility, this would further strengthen the case for involvement of orfc in male sterility.

Exactly how such tandem repeats arise, or are lost and their function in plant mtDNA recombination remains a matter for debate. In yeast, perfect direct repeats found at the sites of excision involved in petite genome formation have been considered to be strong evidence of site-specific recombination events (de Zamaroczy et al. 1983). Currently, our knowledge of plant mtDNA recombination extends only to descriptions of the products of either homologous or site-specific exchanges within 'master' genomic circles or between subgenomic molecules. We are still ignorant of the mechanisms which underlie the recombination processes. Of particular interest is the nature and specificity of the enzymatic complexes or 'recombinases' which function in recombination.

In higher plants no nuclear or mitochondrial gene products have yet been identified which stimulate mtDNA recombination. In yeast, the product of a nuclear PIF gene stimulates recombination in a class of petite mutants, (PIF-dependant mutants), possibly through stabilisation of some mtDNA secondary structure involved in initiating the recombination process (Foury and van Dyck 1985). An accurate determination of the sequences and sequence topologies which are involved in plant mtDNA rearrangements will await the development of in vitro systems for assessing the ability of mtDNA sequences to undergo recombination. Site-directed mutagenesis could be used for testing how recombination is influenced by alterations to mtDNA primary sequence or predicted secondary structures for clues as to how particular mtDNA sequences may act as 'cis' elements in site-specific recombination mediated by interaction with a 'trans-acting' recombinase protein or complex.

7.5 The 'Plasmid-like DNA' in Restorer Sunflower May be an Indication of a Nuclear Influence on mtDNA Replication.

In chapter three a previously unreported '5kb plasmid-like DNA' molecule present in mitochondria from all lines investigated was described. This plasmid-like DNA (hereafter referred to as the '5kb plasmid') is present at low level in fertile, sterile and hybrid sunflower, in contrast to a greatly amplified level found in the restorer line 843R. The plasmid shares homology with the main sunflower mitochondrial genome as detected by hybridisation analysis. Of potentially greater significance is that the 1.2kb EcoRI fragment from sterile sunflower, which hybridises with the 5kb plasmid, represents part of the rearrangement found in this genome. In the absence of sequence evidence, it is tempting to speculate that homology shared between the 5kb plasmid and the
1.2kb EcoRI fragment from sterile sunflower, is due to the presence of a copy, or several copies of the 300bp repeat implicated in the generation of the rearrangement in the sterile genome (section 3.16).

It is interesting to note that the mitochondrial genome organisation of the restorer, sterile and hybrid lines appears to be identical suggesting that the plasmid has not arisen by simple excision from the main genome. It may represent a small subgenomic molecule, present at low copy number in most lines, but which has become specifically amplified in the restorer line due to modification of the origin of replication leading to an increased efficiency of replication.

The difference in abundance of the 5kb plasmid between the hybrid and restorer lines is intriguing if not confusing. The hybrid contains a nuclear genome resulting from a combination of the nucleus from the restorer line and the 

H.annuus nucleus of the sterile line. Thus, one interpretation is that an alteration to the nuclear-mitochondrial combination appears to result in the specific amplification of the plasmid in the restorer sunflower, presumably through an influence on the rate of replication of the molecule relative to the main genome.

7.6 Synthesis of Variant Mitochondrial Proteins in Sunflower

The finding of novel coding regions within the mitochondrial genome of sterile sunflower and the proposal that they may be important phenotypically has been further strengthened by the discovery of synthesis of a variant polypeptide by mitochondria isolated from the CMS cytoplasm. A ca. 15kDa protein, synthesised by mitochondria from the sterile line 842A, correlates with the presence of a variant atpA 3' region in the mitochondrial genome that includes sequences for two ORFs (orf3 and orfd). Orf3 contains 90 codons and the predicted polypeptide product has a molecular weight of only 10.2kDa would thus appear to be too small to encode the variant protein which has a size, estimated by SDS electrophoresis, of ca. 15kDa. However, this estimate may not be an accurate reflection of the molecular weight of the variant protein as a rate of migration in SDS gels can be influenced by the amino acid composition.

A second novel ORF, orfd, was also found within the variant atpA 3' region of mtDNA from sterile sunflower although only part of the 3' sequence of the ORF, encoding 127 codons of the carboxy-terminus of the predicted protein, was obtained. This orfd potentially encodes a protein of at least 14.4kDa and thus is a stronger candidate for the ORF encoding the ca. 15kDa polypeptide synthesised in the sterile mitochondria. Computer searches with the orfd sequence were inconclusive in assigning an identity to the predicted protein it encodes and thus make it impossible to describe a possible function, if indeed it is expressed in mitochondria.
7.7 Molecular Events in the Cytoplasm During Pollen Formation

The majority of the work described in this thesis relates to observations made with mtDNA from etiolated sunflower seedling tissue. Chapter five presents the results of experiments aimed at addressing the tissue specific expression of the CMS trait during the reproductive stages of development which lead to pollen formation.

Most of the available information on cellular changes leading up to and during reproductive development in higher plants come from descriptions based on studies with the light and electron microscope. Havelange et al. (1974) have reported their observations from an analysis of the transition to flowering in the shoot apical meristems of Sinapis alba (mustard) which have been induced to flower (i.e. 'evoked') by long-day photoperiod treatment. They found that the early phase of the morphogenetic transition from vegetative to reproductive state involves a rapid increase in membrane biosynthesis. Tonoplast membrane synthesis is associated with a rise in the number of vacuoles seen in cell sections of 'evoked' apices compared to vegetative meristems. An increase in the total cross-sectional area of mitochondria, was also taken to indicate that de novo synthesis of mitochondrial membranes occurs at this stage of development.

The increase in mitochondria is significant after 18 hours of the start of the induction period and reaches a peak within 62 hours. The actual number of mitochondrial profiles seen in cross-section was also found to increase, and by 62 hours is approximately three times higher in the 'evoked' meristems than in the control vegetative meristems. This suggests that the increase in mitochondrial membrane area results from a multiplication of mitochondria although the exact number of the mitochondria was not determined and an increase in the size of existing mitochondria could not be discounted. An increase in succinate dehydrogenase activity was also found to slightly precede the peak noted for the total mitochondrial membrane area in 'evoked' meristems (occurring at 54 hours), providing further evidence that induction of the reproductive phase involves a rise in cellular respiratory activity.

There are relatively few reports which compare mitochondrial genome level and expression in reproductive tissue with those of vegetative stages of plant development, mainly due to the inaccessibility of tissues such as anthers and pistils in many species. However, in maize, tissue-specific variation in mtDNA levels have been reported (Abbott et al. 1985). The relative levels of the nuclear and mitochondrial genomes were found to differ between tissues (anthers and leaves were examined) and also between anthers from fertile and CMS-T sterile maize, suggesting some alteration to the nucleo-cytoplasmic regulation of mitochondrial genome level may occur in the sterile line. More recently, Mache et al. (1989) have reported that in normal fertile maize the mitochondrial
genome is amplified to reach a maximum level at a specific stage during microsporogenesis. This amplification is associated with increased transcription of the \textit{atp6}, \textit{atp9} and \textit{cob} genes.

The results obtained with sunflower did not indicate that an increase in mtDNA relative to total cellular DNA occurs during the stages of fertile flower development which were examined. It may be that less dramatic changes in mitochondrial genome copy number occur in sunflower reproductive tissue compared to maize and were not detected by the DNA hybridisation analysis. Alternatively, a stage comparable to that of mitochondrial genome amplification during maize microsporogenesis may occur at an earlier stage in sunflower than the youngest stage investigated during this study.

An examination of transcription of the mitochondrial genome has provided preliminary evidence of variation between fertile and sterile sunflower during reproductive development. The level of an unassigned 2.6kb transcript, (partially encoded by the orf73 sequence described in Chapter four) appears to be highly abundant in early fertile flower development relative to the transcripts corresponding to three other mitochondrial genes, \textit{atpA}, \textit{cob} and \textit{atp9}. The abundance of all four mitochondrial transcripts decreases in older tissue as a further indication that the stages examined correspond to a 'post-peak' period of mitochondrial transcription. Although little data relating specifically to sterile tissue was obtained, the mitochondrial transcript levels (specifically the 2.6kb transcript) were found to be abnormally low in very young flower buds relative to cytosolic rRNAs. A reduction in the amount of extractable RNA, was found during the later stages of flower development. In the flower tissues of most plants, some reduction of cytosolic RNA occurs around the meiotic prophase stage of microsporogenesis. This may be due to normal rates of RNA degradation and/or increases in the levels of hydrolytic enzymes in the cytoplasm during this phase of reproductive development (Dickinson 1987).
7.8 The Molecular Basis of Cytoplasmic Male Sterility in Sunflower

The results of the analysis of the mitochondrial genome organisation, expression and protein synthesis in fertile and sterile sunflower can be used to propose an outline model for the molecular basis of the male sterile phenotype. This model leads to proposals for future work described in later sections, which it is hoped will eventually allow a causal link to be established between the molecular biological observations and the biochemical and physiological basis for pollen abortion in CMS sunflower.

It is proposed that the genomic rearrangements described in the sterile sunflower have led to changes in the size of the transcript encoding the α-subunit (ATPA) protein: a 1.9kb transcript in fertile sunflower is replaced by a 3kb transcript in sterile sunflower. The rearrangements have also resulted in the creation of two novel open reading frames (orfC and orfD), one of which may be related to the variant 1.7kb transcript found in mitochondria from sterile sunflower. This 1.7kb transcript may encode the variant ca. 15kDa polypeptide which is synthesised by mitochondria isolated from the sterile, hybrid and restorer lines, but not by those from the fertile (see Figure 7.1).

It is further proposed that the 15kDa polypeptide will be associated, in some way, with the inner mitochondrial membrane where it may impair mitochondrial function and/or biogenesis to a limited extent. This effect may have no consequence for cellular function or differentiation during normal vegetative growth. However, during the period of active mitochondrial biogenesis and function associated with tapetal development and microsporogenesis, the presence of the 15kDa polypeptide and associated impairment of mitochondrial function may have a critical effect on the potential of the mitochondria to support the cellular demand for energy and/or specific metabolites thus leading to tapetal malfunction and failure to produce functional pollen.

In the hybrid and restorer cytoplasms, the transcript patterns relating to the region which has undergone rearrangement are identical to that described in the sterile sunflower. In addition, the hybrid and restorer lines synthesize the variant 15kDa polypeptide. If this protein is indeed causally associated with the CMS phenotype, then it can be proposed that the role of the nuclear restorer genes is to compensate for the deleterious effect on mitochondrial function, leading to the restoration of fertility. It must be stated that the data on which the model shown in Figure 7.1 is based was derived entirely from investigations of mitochondrial genome expression in etiolated tissue and that further examination of the anther tissue specifically affected by the CMS mutation is required.
Figure 7.1
Origins and Consequences of Mitochondrial Genome
Reorganisation Associated with the atpA locus in Sunflower

A summary of the current understanding of organisation of the mitochondrial genome in the vicinity of the atpA locus in fertile and sterile sunflower and the associated variant expression is presented. The atpA gene of the fertile genome is transcribed as 1.9kb transcript. As a consequence of the genomic inversion/insertion events downstream of atpA, the mRNA for this gene is expressed as an extended 3kb transcript in sterile sunflower. A second transcript of 1.7kb transcribed from the rearranged region downstream of atpA in the sterile genome, replaces a 1.3kb transcript in the fertile genome.

The 2.4kb transcript corresponding to the cob gene, located entirely within the genome inversion, is identical in the fertile and sterile sunflower. A 2.6kb transcript, encoded by the sequence immediately adjacent to the genome inversion, is also similarly expressed in fertile and sterile sunflower.

The genome rearrangement predicts no effect on the translation products corresponding to the atpA transcripts, or the cob transcripts between the fertile and sterile cytoplasms. Thus, functional F1 ATP synthase α–subunit and apocytochrome subunit b proteins are present in the mitochondria from both lines.

One of the novel open reading frames, present in the downstream region of the atpA gene in sterile sunflower (orfC and orfD) may correspond to the 1.7kb transcript. Whether this 1.7kb transcript encodes the variant 15kDa polypeptide synthesised by mitochondria from this cytoplasm remains to be shown. The 1.3kb transcript present specifically in fertile sunflower may correspond to the orfB sequence immediately 3' to the atpA coding region, or alternatively, the orf70 sequence located further downstream. The former explanation is supported by the fact that sterile sunflower contains the orf70 sequence but does not express the 1.3kb transcript.

The variant polypeptide may be associated with the mitochondrial membrane system, leading to general or specific dysfunction of a respiratory complex(es) which prevents high rates of respiration. Although 'tolerated' during normal vegetative development, mitochondrial malfunction may become critical at microsporogenesis, leading to abnormal pollen development in CMS sunflower.
Figure 7.1

**FERTILE**

5' 1.9kb

1.9kb atpA transcript

1.3kb transcript

2.4kb cob transcript

2.6kb transcript

**STERILE**

5' 3kb

3kb atpA transcript

1.7kb transcript

2.4kb cob transcript

2.6kb transcript

15kDa variant protein

ATPA protein

COB protein

?
7.9 Proposals for Future Work

Ultimately, only a combination of molecular, biochemical and cytological investigations of mitochondrial genome expression and metabolic activity during microsporogenesis will provide a complete understanding of the cause of male sterility in sunflower and other species. The work presented in this thesis, has been, of necessity, descriptive. At the inception of this project nothing was known of expressed variation between the mitochondrial genomes of fertile and sterile sunflower. In the simplest terms, the results which have been obtained are correlations. The difficult aspect in an investigation of CMS is not necessarily proving the existence of molecular variation in the mitochondria which correlates with the sterile phenotype, but is the subsequent elucidation of the mechanism by which mitochondrial malfunction arises as a consequence. Several specific proposals for continuation of the investigation of CMS in the sunflower system are outlined below. These are based on the information that we now have and will hopefully lead to a complete elucidation, from molecular mutation to phenotypic aberration, of how the male sterile phenotype is manifest in sunflower.

7.9.1 Sequence and Transcript Mapping

As one of the contributing regions to the rearrangement present in the genome of sterile sunflower, sequencing of the inversion 'endpoint' located within the 0.8kb EcoRI fragment from the fertile line is required. This will provide an accurate description of the size of the 200-300bp repeat potentially involved in this genome reorganisation. During the sequence analysis, M13 clones and oligonucleotides corresponding to specific regions of the \textit{atpA} loci in fertile and sterile sunflower were produced which can be used to further investigate the transcripts homologous to these regions of the genomes.

In particular the sequences which encode the 1.7kb transcript in sterile sunflower, and the 1.3kb transcript in fertile sunflower, should be defined by the use of S1 and primer extension experiments. The localisation of the 5' and 3' termini of these transcripts, within the \textit{atpA} 3' region from which they derive, will indicate whether the variant 1.7kb transcript encodes a 'modified' (orfc) or a 'novel' (orfd) product. In addition the 5' and 3' termini of the 1.9kb \textit{atpA} mRNA from fertile sunflower and the 3kb \textit{atpA} transcript of sterile sunflower can also be mapped to define the sequences involved in initiation and termination.

Whether the 5kb plasmid, present at an amplified level in restorer sunflower, is in any way associated with suppression of the CMS phenotype remains to be shown. This molecule should be isolated and cloned for DNA sequence analysis to provide evidence of any potential coding function that
may be associated with the plasmid. Sequencing will also determine the extent of homology within the plasmid to the main mitochondrial genome and could lead to a description of the formative recombination events which led to its origin.

7.9.2 Anther Development and Transcript Analysis

Mitochondrial transcription during reproductive development in fertile and sterile sunflower should be re-examined using a range of mitochondrial and nuclear gene probes. Ideally, this should be done with anther tissue from the earliest possible differentiated state in both fertile and sterile lines and compared with several different vegetative tissues. In this way a quantitative description of the relative levels of the mitochondrial genomes, at several stages of development, between phenotypes will be possible.

It may then be feasible to construct mitochondrial cDNA libraries when sufficient amounts of mtRNA from fertile and sterile sunflower anthers are available. Cross-hybridisation with labelled mtRNA isolated from fertile and sterile anthers could be used to screen such libraries, with the aim of identifying a region or regions of the mitochondrial genome which are differentially transcribed in anther tissue from the two phenotypes. Differential screening with mtRNA from non-reproductive tissue (e.g. etiolated seedling or leaf tissue) would be necessary to indicate a developmental correlation. Ideally, in situ hybridisation experiments should be performed with probes derived from the rearranged loci found in mtDNA of sterile sunflower, in order to determine whether differential expression of sequences encoded by this region occurs in tapetal tissue.

7.9.3 Mitochondrial Protein Purification and Functional Characterisation of the Variant 15kDa Polypeptide

The location of the variant 15kDa protein synthesised by mitochondria from the sterile cytoplasm of sunflower and its function, if any, remain to be determined. Purification of a sufficient quantity of this protein followed by amino acid sequence analysis, would resolve which, if either, of the orfc and orfd sequences encodes it. In addition, the use of non-denaturing gels could indicate an association the variant protein may have with a complex or complexes of the mitochondrial membrane system. However, as has been indicated by attempts to assign a functional role to the 13kDa protein in CMS-T maize, interpretations based on in organello translation can be misleading (Leaver et al. 1988). Non-specific aggregation between unassembled translation products may occur under the conditions used in vitro which are not representative of the situation in the living plant.
Notwithstanding technical problems in the purification of the ca. 15kDa polypeptide from sterile sunflower it should be possible to raise specific antibodies with which to examine whether the level of synthesis of this variant protein is altered in different tissues, and in particular for evidence of increased synthesis in anthers. The availability of such antibodies will allow an investigation of whether a modified version of the 15kDa protein or an altered association of the protein with other mitochondrial components occurs in those plants which have been restored to fertility.

A quantitative comparative assessment of the levels of specific mitochondrial proteins, using the technique of Western blotting and antisera raised against particular proteins (e.g. $\alpha$-subunit, COXI and the 15kDa polypeptide), in various tissues of fertile and sterile sunflower could also be made. By comparing these data to those of transcript analysis and in organello protein synthesis, it may be possible to determine to what extent translational controls of mitochondrial gene expression occur during different stages of development. Specifically in relation to anther differentiation, an analysis of whether different genes are transcriptionally and/or translationally regulated may prove invaluable in an assessment of altered gene expression relating to CMS.

Preliminary experiments suggest that the use of different substrate and energy sources can influence the relative rates of synthesis and/or stability of particular proteins within mitochondria, in vitro as was described in Chapter six. This may be an indication of the in vivo situation where substrate supply, mediated by transport into mitochondria and/or affects on the TCA cycle, could play an important role in influencing the expression of mitochondrial proteins in plant development. Differential translation of mitochondrial proteins at different stages of plant development have been described for maize (Newton and Walbot 1985). However, only in Petunia has there been any report of altered mitochondrial protein synthesis specifically in reproductive tissue (Young and Hanson 1987).

7.9.4 Biochemical Characterisation of Sunflower Mitochondria

In Chapter six, the results of preliminary experiments examining the effect of energy sources on mitochondrial protein synthesis were described. These have provided some circumstantial evidence of abnormal biochemical function in mitochondria from sterile sunflower indicated by their apparent inability to utilise succinate as an efficient energy source. It is of interest that mitochondria isolated from CMS-T sterile conferring cytoplasm in maize have also been found to show low levels of protein synthesis with succinate as the supplied energy source (A.D. Liddell, personal communication). There is no evidence for a specific association of the 13kDa variant protein found in this
cytoplasm with complex II of the inner mitochondrial membrane. Thus a speculative interpretation is that complex II may be more susceptible to any alterations to the surrounding 'molecular environment' caused by the presence of variant polypeptides such as the 13kDa in CMS-T maize and the 15kDa in sunflower. By use of an oxygen electrode, respiration in mitochondria isolated from fertile and sterile sunflower could be compared for evidence of biochemical variation between the phenotypes. The rates of respiration and the degree of coupling in mitochondria, in the presence of succinate as an energy source, should be examined with mitochondria isolated from the two lines of sunflower in order to clarify the results from the in organello translation experiments.

To date, there are very few studies which address changes in cellular energy specifically associated with anther development as an indication of the importance of mitochondrial function at this time. Recently, the use of high performance liquid chromatography (HPLC), to compare anther extracts from fertile and CMS Petunia has shown that dramatic differences in the levels of adenine and pyridine nucleotides are present in the two cytoplasms (Liu and Dickinson 1989). Cytochemical localisation techniques show that in normal anthers an increase in NADPH occurs specifically in the tapetal cell layer and reaches a peak, in the anthers, at the early prophase of pollen meiosis. Anthers from CMS plants at approximately the same stage of development do not show an accumulation of NADPH. An early precursor for the synthesis of nucleic acids a lack of NADPH in the tapetal cells of sterile Petunia could explain the abnormally low levels of DNA synthesis noted in this tissue (Liu et al. 1987).

7.10 The Changing Role of Mitochondria During Plant Development.

In higher plants, changes in mitochondrial activity provide energy and metabolites for a wide range of developmental transitions. The onset of flowering is just one example of a stage of active mitochondrial biogenesis, fruit ripening and early seedling development are others. Important questions to be addressed are when does biogenesis of mitochondria occur and how are the processes of membrane biosynthesis and respiratory function regulated?

It has been shown that active mitochondrial biogenesis may be very stage specific during higher plant development. For example, in a developing wheat leaf, mitochondrial biogenesis is restricted to the extreme basal sections of the leaf (Topping 1987). Even when net active mitochondrial biogenesis is not occurring, it is assumed that turnover of mitochondrial protein does occur, and that there is continual import of nuclearly synthesised proteins to maintain respiratory and synthetic function. Quantitative and qualitative variation in the proteins which are imported into mitochondria may represent an important level of control in mitochondrial biogenesis and function by
modulating the biochemical environment, and hence the metabolic potential, within the organelle. The composition of the cytosol itself may act at various levels to coordinate mitochondrial activity with the energy requirements of the cell. Feedback mechanisms, mediated by exchange of ion and metabolic intermediates, may relay 'information' between the mitochondria and nucleus, ranging from a general build up of precursors to requirements for particular mitochondrial proteins and activation of specific membrane receptors.

Only by investigating the regulation of expression of mitochondrial genes and nuclearly encoded proteins which are targeted to mitochondria in differentiating systems, can we begin to explain how alterations in mitochondrial composition and function contribute to changing energy demands in plant development. A detailed understanding will, in part, be facilitated by the use of molecular and *in vitro* culture techniques to manipulate the expression of particular genes which function in mitochondria.

### 7.11 Genetic Engineering of Plant Mitochondria—Some Future Prospects

A major aim of studying mitochondrial function in higher plants is to determine ways of directly modifying the efficiency of energy conversion and metabolic activity to produce phenotypic effects. The ability to genetically transform plant cells allows the introduction of defined DNA sequences to create new, inherited genetic loci. By regeneration of whole organisms from single transformed cells, the consequences of specific gene manipulation for the phenotypic alterations in transgenic plants can be assessed. Changes of the structure and/or copy number of genes encoding mitochondrial proteins corresponding to specific metabolic processes, or transport mechanisms, could potentially influence energy economy in plants cells and lead to improvements in the yields of important crop species. To be useful, molecular engineering of mitochondrial performance has to be directed and requires a detailed understanding of those processes which regulate mitochondrial gene expression *in vivo*.

Nuclear transformation with genes incorporating suitable sequences for mitochondrial targeting of the protein product represents one way of introducing polypeptides, not normally found within mitochondria, into the organelle. To date, two bacterial 'reporter' proteins, chloramphenical acetyltransferase (CAT) and β-glucoronidase (GUS) have been specifically 'addressed' to tobacco mitochondria using targeting sequences from the pea β-subunit of the F₁ATPase and a yeast mitochondrial tRNA synthetase respectively (Boutry *et al.* 1987, Schmitz and Lonsdale 1989). These types of experiments are important in defining the specific sequences required for efficient import of proteins which must be considered in the production of artificial gene constructs designed to target functional proteins to mitochondria.
Genetic transformation of the plant mitochondrial genome has not yet been achieved. De Block et al. (1985) claimed to have transformed tobacco chloroplasts with the bacterial CAT gene, via Agrobacterium tumefaciens mediated transfer. The use of the 'Ti-plasmid', contained within this naturally occurring soil bacterium, as a transformation vector exploits its ability to accommodate the insertion of additional sequence and faithfully transfer it into the plant genome upon infection (Hernalsteens et al. 1980). Either the 'Ti-plasmid' transfer system or direct high velocity delivery of DNA coated microprojectiles into mitochondria (Klein et al. 1987) offer the best prospects for achieving mitochondrial transformation in the future. The recent report that the 11.3kb linear plasmid found in mitochondria of Brassica napus (rapeseed) can be transmitted to progeny, via the paternal line during sexual reproduction, is also of potential interest with regard to mitochondrial transformation (Erickson et al. 1989). It would appear that in B.napus, there is at least some transfer of paternal mitochondria from the pollen to the zygote during fertilisation. This would arise if the sperm cell which fuses with the egg cell donates some mitochondria to the resulting embryo. The contribution could be very minor, but provide sufficient influx of paternal mitochondria to allow transfer of the plasmid. If this plasmid could be modified in vitro, pollen mediated transfer could provide an efficient method for introducing this plasmid to other lines as a means of achieving mitochondrial transformation.

Transformation of nuclear and when it becomes a reality mitochondrial genomes with multicopies of genes for mitochondrial proteins under the control of strong promoters may be a possible method for analysing feedback controls that coordinate expression of the two genetic systems. It may be possible to determine whether overproduction of individual proteins has specific effects or influences particular stages of plant development. By directed mutagenesis of selected genes, to cause changes in rates of transcription, translation, and assembly into complexes of particular proteins, the consequences for expression of other proteins and for mitochondrial activity can be evaluated.

The ability to reintroduce defined DNA segments into plant cells has allowed the identification of sequences which modulate plant gene transcription during changes in the environmental conditions and within specific tissues. Regulatory sequences and the corresponding 'trans-acting' DNA binding factors, which mediate differential expression, are being characterised in a number of different systems. For example, the regulatory signals or light responsive elements (LREs) involved in modulating chloroplast gene expression with changes in the level and quality of light have been extensively investigated (Cuozzo et al. 1988, Deng et al. 1989).
The isolation of genes corresponding to mRNAs expressed specifically at stages of plant development as diverse as seed germination and fruit ripening is being used to define sequences which control tissue specific expression. Of particular interest with regard to CMS are the isolation of cDNA clones encoding mRNAs which are specifically expressed in anther tissue (Gasser et al. 1988, Twell et al. 1989). In vitro translation of polyA+ RNA extracted from several different reproductive organs of tomato have been used to indicate that there are significant quantitative and qualitative differences in the nuclear mRNA of immature and mature anthers (Gasser et al. 1988). In situ hybridisation techniques are providing elegant demonstrations of the localisation of anther mRNAs within specific cell layers of the tapetal and surrounding tissues. These cDNAs can be used to identify the corresponding genomic clones for characterisation of the 5' and 3' flanking regions to learn more of how their expression may be regulated. The prospect that such elements could be used to regulate specific mitochondrial genes is unlikely to be realised for some time although it is an attractive thought with regard to the manipulation of CMS.

Once the mechanism of CMS has been fully elucidated in species such as maize, Petunia or sunflower, it may be possible to engineer CMS in those plants which do not normally show the trait by basing it on one of the naturally occurring systems. This will depend upon the availability of anther specific promoters to mimic the natural manifestation of the male sterile phenotype and to avoid uncontrolled expression of potentially lethal proteins. The use of antisense RNA may provide a method for regulating the expression of introduced 'CMS' genes and for assessing their efficiency.

An inherent assumption in engineering CMS is that artificially introducing a gene characterised as causal in CMS in one species will induce the same phenotype in another. However this may not be the case; in maize and Petunia the implicated variant polypeptides are of different amino acid composition and structure. The same may be true of the 15kDa of sterile sunflower suggesting these variant polypeptides are likely to influence different aspects of mitochondrial function. Conversely, slight differences in mitochondrial membrane composition between species may cause variations in the function of variant polypeptides. It is, of course, imperative that sources of engineered CMS are both stable under varying environmental conditions and also capable of being controllably restored to fertility. The economic consequences of discovering genetic linkage of the 'T-toxin' sensitivity with CMS-T sterility is a further reason that extensive assessment by field trials will always be required for any artificially induced source of male-sterility.
7.12 Conclusion

We have only a very limited understanding of how CMS operates in flowering plants despite its extensive use in the breeding of several important crop species. There is considerable evidence to implicate mtDNA lesions and mitochondrial malfunction as the key factors in abnormal pollen development. However, important questions concerning the mechanism by which altered mitochondrial expression becomes 'lethal' specifically within anther tissue remain to be answered. Expanding our knowledge of changes in mitochondrial biochemistry which contribute to normal reproductive development is crucial to an understanding of the earliest malfunction which characterises CMS plants. Discovering the molecular and cellular basis of the phenotype is essential to the development of systems for the artificial induction of male-sterility. In the process, there is much to be gained from an understanding of the mechanisms which control mitochondrial genome organisation and how this influences genome expression within the organelle. The commercially attractive prospect of a complete understanding of the CMS trait, in even a single species, superimposes on the intriguing challenge of explaining the origin and function of the complexity of the mitochondrial genomes of higher plants.


Benne R., van den Burg J., Brakenhoff J.P.J., Sloof P., Van Boom J.H. and Tromp M.C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819-826


Birnboim H.C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7, 1513-1523


Dawson A.J., Jones V.P. and Leaver C.J. (1984) The apocytochrome b gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding site. *EMBO J.* 3, 2107-2113


213


Dujon B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S gene of yeast strains having different alleles at the o and rib-1 loci. Cell 20, 185-197

Eilers M., Oppliger W. and Schatz G. (1987) Both ATP and an energized membrane are required to import a purified precursor protein into mitochondria. EMBO J. 6, 1073-1077


Folkerts O. and Hanson M.R (1989) Three copies of a single recombination repeat occur on the 443kb master circle of the *Petunia hybrida* 3704 mitochondrial genome *Nucl. Acids Res.* 17, 7345-7357


Glab N. (1988) Yeast mitochondria as a model system for the study of cytoplasmic male sterility in higher plants. Abstracts from the 3rd International Workshop on The Mitochondrial Genome of Higher Plants


Hansen B.M. and Marker K.A. DNA sequence and transcription of a DNA minicircle isolated from male-fertile sugar beet mitochondria. *Nucl. Acids Res.* 4747-4756


Isaac P.G., Jones V.P. and Leaver C.J. (1985b) The maize cytochrome c oxidase subunit I gene: sequence, expression and rearrangement in cytoplasmic male sterile plants. EMBO J. 4, 1617-1623


Joyce P.M.M, Spencer D.F., Bonen L. and Gray M.W. (1988) Genes for tRNAP^{10}, tRNA ^{1Y}, and two tRNAs^{Ser} in wheat mitochondrial DNA. Plant Mol. Biol. 10, 251-262


Leroy P., Bazetoux S., Quetier F., Delbut J., Bervillé A. (1985) A comparison between mitochondrial DNA of an isogenic male-sterile (S) and male-fertile (F) couple (HA89) of sunflower. *Curr. Genet.* 9, 245-251

222


Mache R., Monéger F., Niogret M.-F, Mandaron P. and Barakate A. (1989) Analysis of organelle DNA and expression of some of their genes during male and female flower development in maize. Abstracts from the EMBO-EMBL Symposium on *Molecular Communication in Higher Plants*


Manson J.C., Liddell A.D., Leaver C.J. and Murray K. (1986) A protein specific to mitochondria from S-type male-sterile cytoplasm of maize is encoded by an episomal DNA. *EMBO J.* 5, 2775-2780


228


Schuster W. and Brennicke A. (1987a) Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera*: is genetic information transferred between organelles via RNA? *EMBO J.* 6, 2857-2863

Schuster W. and Brennicke A. (1987b) Interorganellar sequence transfer: plant mitochondrial DNA is nuclear, is plastid, is mitochondrial. *Plant Sci.* 54, 1-10


Small I.D., Isaac P.G. and Leaver C.J. (1987) Stoichiometric differences in DNA molecules containing the atpA gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *EMBO J.* 6, 865-869


234


