Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
TRANSFORMATION AND MARKER GENE REMOVAL STRATEGIES FOR THE PLASTID GENOME OF _BRASSICA NAPUS_

Seamus Coyne

A thesis submitted to the University of Dublin for the degree of Doctor of Philosophy

Smurfit Institute of Genetics,
University of Dublin,
Trinity College,
Dublin 2,
Ireland.  

July, 2002
DECLARATION

This thesis has not been previously submitted to this or any other University for examination for a higher degree. The work presented here is entirely my own. This thesis may be made available for consultation within the University library. It may be photocopied or loaned to other libraries for the purposes of consultation.

Seamus Coyne
ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Tony Kavanagh for his help and guidance throughout the duration of my PhD. Thanks to Nga for her help and wonderful patience. Thanks also to Alan, Paul, Sinead and Jim for their friendship and help. I wish them all the best in their future careers and lives.

I would like to express my sincere gratitude to my parents Seamus and Celestine and to Sinead, Dorothy and Gerard for their constant support and encouragement. Thanks also to Pauline and all my friends both in Dublin and back home.
SUMMARY

Despite the fact that the first transplastomic higher plant was produced more than a decade ago, transformation of the plastid genome and the regeneration of homoplasmic, fertile plants has so far been achieved in only 2 species (*Nicotiana tabacum* (tobacco) and *Lycopersicon esculentum* (tomato)) and the technology is routinely successful in only a single species, tobacco. The application of this important technology to the major, agronomically important crop species is therefore a major challenge. The developments described in this thesis address some of the important issues surrounding the current state of plastid transformation technologies and describes the development of tools for plastid transformation in *Brassica napus* (oilseed rape).

As a starting point for these investigations, a molecular characterization of the *B. napus* plastid genome was undertaken. This involved the generation of restriction fragment clone libraries, the identification of overlapping clones and the assembly of a physical map of the genome relative to that of *Arabidopsis thaliana*. Coverage of 97% of the *B. napus* plastid genome was achieved using this strategy and in addition, a total of 56.6kb of nucleotide sequence (approximately half of the unique sequence of the genome) was determined. The resulting data were used to investigate phylogenetic relationships between *B. napus* and 7 other species for which complete plastid genome sequences were available, and to facilitate the construction of plastid transformation vectors. This analysis confirmed the relatively recent ancestry that exists between the species *A. thaliana* and *B. napus*, with the divergence date estimated to be approximately 14.5-21.7 million years ago. With sequence similarities being so high, *Brassica* biotechnology (both nuclear and plastid transgenics) will surely benefit from the availability of the entire *Arabidopsis* DNA sequence. The divergence dates between the other species in the study were also estimated and are in agreement with the theory that a rapid angiosperm species radiation occurred approximately 100 million years ago. The calculation of nucleotide substitution rates using Kimura’s method allowed for a direct comparison to be made between the coding sequence and non-coding sequences in the plastid genome. They revealed that selected intergenic regions
were subject to higher mutational rates, thus confirming their suitability as targets for foreign gene integration.

Two types of *Brassica*-specific chloroplast transformation vector were constructed. The first, pPC1, contains sequences spanning the ribosomal RNA operon (rrn) and the gene coding for ribosomal protein S12 (rps12 (3')). The vector was designed to direct the insertion of transgenes by homologous recombination into the *trnV-rps12* 3' intergenic spacer in the inverted repeat (IR) region of the *B. napus* plastid genome. Selection for pPC1-mediated transformation was based on engineered point mutations in the 16S rRNA and *rps12* (3') genes that confer spectinomycin and streptomycin “binding-type” resistance, respectively. In contrast, the second vector, pZBl was designed to direct transgene insertion into the intergenic spacer between the *rbcl* and *accD* genes. Selection in this vector was based on detoxification of streptomycin and spectinomycin by the bacterial antibiotic resistance gene aadA, which encodes amino-glycoside 3' adenylation transferase.

Unfortunately, due to the lack of an established protocol for chloroplast transformation in *Brassica napus*, the transformation vectors pPC1 and pZBl were not tried in this species. Instead, pPC1 was tested by biolistic bombardment into tobacco leaf tissue. In spite of the host-vector sequence heterogeneity, two independent spectinomycin resistant plantlets were regenerated that both contained the correct vector-specific antibiotic resistance marker. Neither plantlet contained the streptomycin resistance marker or the foreign gene multiple cloning site. This was postulated to be due to the lack of localised homogeneity in the *rps12* (3') gene compared to the *rrn16S* gene, which was highly conserved (99.5%), even between two species as divergent as tobacco and *B. napus*.

Finally, in response to the current concerns over the use of antibiotic resistance genes in transgenic plants, the development of a marker gene removal strategy for plastids, based on the bacteriophage P1 Cre/loxP system, was also investigated. Transgenic tobacco plants were produced that expressed a plastid-targeted Cre recombinase fusion protein from a transgene located in the nucleus.
The resulting Cre recombinase protein was shown to be targeted to chloroplasts by means of its N-terminal transit peptide domain and to be functional in an *in vitro* marker gene excision assay based on the activation of GFP fluorescence.
ABBREVIATIONS

A   adenine
aadA  adenosyl-3'-adenyltransferase gene
Ac  accession number
accD  acetyl coenzyme A carboxylase gene
Acetyl CoA  acetyl coenzyme A
AGF  AAG binding factor
AIDS  anti immune deficiency syndrome
Ala  alanine
ATP  adenosine-5'-triphosphate
atp  ATPase subunit genes
BA  betaine aldehyde
BADH  betaine aldehyde dehydrogenase
BAP  benzylaminopurine
bar  glufosinate resistance gene
BLAST  basic local alignment search tool
bp  base pairs
BSA  bovine serum albumin
Bt  Bacillus thuringiensis
C  cytosine
°C  degrees celsius
CaCl₂  calcium chloride
cemA  chloroplast envelope protein
clpP  chloroplast ATP-dependent protease gene
cm  centimetre
codA  cytosine deaminase gene
cpDNA  chloroplast DNA
cre  cre recombinase gene
cry  crystal toxin protein gene
cv  cultivar
dATP  2', 3'-deoxy-adenosine-5'-triphosphate
dCTP  2', 3'-deoxy-cytidine-5'-triphosphate
dGTP  2', 3'-deoxy-guanosine-5'-triphosphate
dNA  deoxyribonucleic acid
dTTP  2', 3'-deoxy-thymidine-5'-triphosphate
E. coli  Escherichia coli
EDTA  ethylenediamine tetraacetic acid
EPSPS  5-enol-pyruvylshikimate-3-phosphate synthase
g  gravitational acceleration / gram
G  guanine
GFP  green fluorescent protein
Gly  glycine
GM  genetically modified
GUS  β-glucuronide
HCl  hydrochloric acid
He  helium
HEPES  N-[2-Hydroxyethyl]piperazine-N'-[2-ethansulfonic acid]
Hg  mercury
hST  human somatotropin
Ig immunoglobulin
in inch
infA chloroplast initiation factor 1 gene
IPTG isopropyl-β-D-thiogalactopyranoside
IR inverted repeat
kb kilobase
kBq kilobecquerel
KCl potassium chloride
kDa kilodalton
kV kilovolt
L litre
LB Luria-Bertani
LSC large single copy
μF microfarad
μg microgram
μl microlitre
μM micromolar
M molar
mA milliampere
MBP maltose binding protein
met methionine
mg milligram
Mg²⁺ magnesium ion
MgCl₂ magnesium chloride
MgSO₄ magnesium sulphate
min minute
ml millilitre
mm millimetre
mM millimolar
mRNA messenger RNA
MS Murashige and Skoog medium
Myr million years
NAA α-naphthalene acetic acid
NaCl sodium chloride
NADH β-nicotinamide adenine dinucleotide, reduced form
NaOH sodium hydroxide
NBM MS medium with hormones and vitamins
NCBI national centre for biotechnology information
ndh chloroplast NDH complex gene
NEP nucleus encoded plastid RNAP
ng nanogram
(NH₄)₂SO₄ ammonium sulphate
nm nanometre
npt neomycin phosphotransferase gene
OD optical density
oligo oligonucleotide
ORF open reading frame
oz ounce
PAGE polyacrylamide gel electrophoresis
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>uidA</em></td>
<td>β-glucuronide gene</td>
</tr>
<tr>
<td><em>UTR</em></td>
<td>untranslated region</td>
</tr>
<tr>
<td><em>UV</em></td>
<td>ultra-violet</td>
</tr>
<tr>
<td><em>WT</em></td>
<td>wild-type</td>
</tr>
<tr>
<td><em>X-Gal</em></td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
<tr>
<td><em>ycf</em></td>
<td>hypothetical chloroplast open reading frame</td>
</tr>
<tr>
<td>Section</td>
<td>Page No.</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>GENERAL INTRODUCTION</strong></td>
<td></td>
</tr>
<tr>
<td><strong>1.1. The Chloroplast</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1.1. The Chloroplast Genome</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2. Chloroplast Genome Structure</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3. Chloroplast Genes</td>
<td>3</td>
</tr>
<tr>
<td>1.1.3.1. Genes Encoding Polypeptides of the Photosynthetic System</td>
<td>3</td>
</tr>
<tr>
<td>1.1.3.2. Genes Encoding Polypeptides of the Genetic System</td>
<td>4</td>
</tr>
<tr>
<td>1.1.3.3. Genes coding for the tRNAs and rRNA</td>
<td>4</td>
</tr>
<tr>
<td><strong>1.2. Chloroplast Gene Expression</strong></td>
<td>5</td>
</tr>
<tr>
<td>1.2.1. Chloroplast Transcription</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1.1. The Plastid-Encoded RNA polymerase (PEP)</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1.1.1. PEP Promoters</td>
<td>6</td>
</tr>
<tr>
<td>1.2.1.1.2. Sigma (σ) Factors in Plastid Transcription</td>
<td>6</td>
</tr>
<tr>
<td>1.2.1.2. Nuclear-Encoded RNA Polymerase (NEP)</td>
<td>7</td>
</tr>
<tr>
<td>1.2.1.2.1. NEP Promoters</td>
<td>8</td>
</tr>
<tr>
<td>1.2.1.3. Transcription of the rRNA Operon</td>
<td>8</td>
</tr>
<tr>
<td>1.2.2. Chloroplast mRNA Processing</td>
<td>9</td>
</tr>
<tr>
<td>1.2.3. Chloroplast Translation</td>
<td>10</td>
</tr>
<tr>
<td><strong>1.3. Chloroplast Genome Evolution</strong></td>
<td>12</td>
</tr>
<tr>
<td>1.3.1. Chloroplast Origins</td>
<td>12</td>
</tr>
<tr>
<td>1.3.2. Gene Transfer to the Nucleus</td>
<td>14</td>
</tr>
<tr>
<td><strong>1.4. Chloroplast Transformation</strong></td>
<td>16</td>
</tr>
<tr>
<td>1.4.1. DNA delivery into Chloroplasts</td>
<td>16</td>
</tr>
<tr>
<td>1.4.1.1. Biolistic Bombardment</td>
<td>17</td>
</tr>
</tbody>
</table>
1.4.1.2. Polyethylene Glycol Treatment
1.4.1.3. Femtoinjection Technique
1.4.1.4. Future Possibilities in DNA Delivery to the Plastid
1.4.2. Integration of Foreign DNA into the Plastome
1.4.3. Selectable Markers
   1.4.3.1. Chloroplast Mutations conferring Resistance to Antibiotics
   1.4.3.2. Heterologous Gene Markers
   1.4.3.3. Marker Gene Removal from the Chloroplast
1.4.4. Investigating Chloroplast Metabolism via Transformation

1.5. Biotechnological Applications of Plastid Transformation
   1.5.1. Pest Resistance
   1.5.2. Herbicide Tolerance
   1.5.3. Disease Resistance
   1.5.4. Recombinant Protein Production in Plastids
   1.5.5. Production of Edible Vaccines

1.6. Advantages of Plastid vs. Nuclear Transformation
   1.6.1. Transgene Containment
   1.6.2. Polycistronic “Operon” Expression in Plastids
   1.6.3. High-level Expression
   1.6.4. Compartmentalisation of Expression
   1.6.5. Limitations and Future Directions

1.7. Thesis Aims
CHAPTER 2
THE PLASTID GENOME OF BRASSICA NAPUS

2.1. Introduction

2.1.1. Brassica napus (oilseed rape)

2.1.2. Evolution of Angiosperms
   2.1.2.1. Evolution of the Brassica Lineage

2.1.3. Construction of B. napus Plastid Genome Libraries

2.2. Materials and Methods

2.2.1. Materials
   2.2.1.1. Plant Materials
   2.2.1.2. Bacterial Strains and Plasmids
   2.2.1.3. Enzymes
   2.2.1.4. E. coli Growth and Selective Media
   2.2.1.5. Oligonucleotides used for DNA Sequencing

2.2.2. Methods
   2.2.2.1. Molecular Cloning Techniques
   2.2.2.2. Purification of Chloroplast DNA
   2.2.2.3. Preparation of E. coli Competent Cells and Transformation
   2.2.2.4. Purification of Plasmid DNA (Boiling Method)
   2.2.2.5. \(^{32}\)P Labelling of Chloroplast DNA Digests, Plasmids and Isolated DNA Fragments
   2.2.2.6. Transfer of DNA from Agarose Gels to Nitrocellulose Membrane
   2.2.2.7. Screening of Recombinant Bacteria by Colony Blotting
   2.2.2.8. Hybridisation of DNA-Containing Membranes
   2.2.2.9. Sequence Data Assembly and Analysis

2.3. Results

2.3.1. Construction of B. napus Plastid Genome Libraries
2.3.2. Nucleotide Sequencing of the *B. napus* Plastid Genome 56
  2.3.2.1. Phylogenetic Analysis 57
  2.3.2.2. Intergenic Sequence Analysis 58

2.4. Discussion 61
  2.4.1. *B. napus* Plastid Genome Restriction Fragment Libraries 61
  2.4.2. Plastid Genome Sequence Analysis 62
    2.4.2.1. Coding Sequence 62
    2.4.2.2. Phylogenetic Analysis 62
    2.4.2.3. Nucleotide Substitution Rates 65
    2.4.2.4. Intergenic Sequence Analysis 66
CHAPTER 3
CONSTRUCTING CHLOROPOLAST GENOME
TRANSFORMATION VECTORS FOR B. NAPUS

3.1. Introduction
3.1.1. Plastid Transformation Vectors
3.1.2. Plastid Transformation Efficiencies
3.1.3. Plastid Transformation in Higher Plants other than Tobacco

3.2. Materials and Methods
3.2.1. Materials
3.2.1.1. Plasmids
3.2.1.2. In vitro Plant Material
3.2.1.3. Oligonucleotides used for DNA Sequencing
3.2.1.4. Oligonucleotides used for PCR
3.2.1.5. Oligonucleotides used to make Synthetic Polylinkers and Insertions

3.2.2. Methods
3.2.2.1. Overlap Extension PCR
3.2.2.2. Annealing Oligonucleotides
3.2.2.3. ³²P Labelling of Oligonucleotides
3.2.2.4. Plasmid DNA Maxiprep
3.2.2.5. Preparation of Tungsten Particles for Biolistic Transformation
3.2.2.6. Coating Washed Tungsten Particles with DNA
3.2.2.7. Biolistic Transformation of Tobacco Leaf Tissue
3.2.2.8. Regeneration of Stable Transformants

3.3. Results
3.3.1. Construction of B. napus Plastid Transformation Vectors
3.3.1.1. Construction of the Transformation Vector pPC1
3.3.1.1. Introduction of a Spectinomycin Resistance Point Mutation 82
3.3.1.1.2. Introduction of a Multiple Cloning Site 84
3.3.1.1.3. Introduction of a Streptomycin Resistance Point Mutation 84
3.3.1.1.4. Completion of the pPC1 Construct 85
3.3.1.1. Construction of the Transformation Vector pZBl 86

3.3.2. Transformation of the Tobacco Chloroplast Genome using the B. napus-specific vector pPC1 88
3.3.2.1. Molecular Analysis of the Putative pPC1 Transformants 89

3.4. Discussion 91
3.4.1. B. napus Chloroplast Transformation Vectors 91
3.4.2. Tobacco Chloroplast Transformation using pPC1 91
3.4.2.1. Transformation Efficiency of pPC1 in Tobacco 91
3.4.2.2. Universal Transformation Vectors 92
3.4.3. The B. napus-specific vector pZBl 94
CHAPTER 4
A CHLOROPLAST-TARGETTED CRE RECOMBINASE

4.1. Introduction

4.1.1. Site Specific Recombination Systems
4.1.1.1. The Cre Recombinase/loxP system
4.1.1.2. Other Recombinase Systems

4.1.2. Heterologous Expression of Recombinases in Plants
4.1.2.1. Site-specific Gene Insertion using the Cre/loxP System
4.1.2.2. Site-specific DNA Excision from the Nuclear Genome

4.1.3. Developing a Cre/loxP Recombination System for Chloroplasts

4.2. Materials and Methods

4.2.1. Materials
4.2.1.1. Plasmids
4.2.1.2. Antibodies
4.2.1.3. Oligonucleotides used during the Assembly of the Target Construct for the Cre Recombinase Activity Assay
4.2.1.4. Sequencing Primers used to confirm the Correct Assembly of pBAD-lox-aadA-N-GFP
4.2.1.5. Oligonucleotides used during the Assembly of the Cre Recombinase Expression Cassettes pRok-Cre and pRok8-Cre
4.2.1.6. PCR Primers used to Screen for Transgenic Plants Transformed with pRok-Cre and pRok8-Cre
4.2.1.7. Sequencing Primers used to confirm the Correct Assembly of pRok-Cre and pRok8-Cre

4.2.2. Methods
4.2.2.1. Preparation of Electroporation Competent *Agrobacterium* Cells

4.2.2.2. *Agrobacterium* Transformation by Electroporation

4.2.2.3. Transformation of Tobacco

4.2.2.4. Preparation of Tobacco DNA Extracts

4.2.2.5. Preparation of Tobacco Protein Extracts

4.2.2.6. Western Blotting

4.2.2.7. *In vitro* Cre Activity Assays

4.3. Results

4.3.1. Construction of a Cre Excision Assay Plasmid *(pBAD-lox-aadA-N-GFP)*

4.3.1.1. Introduction of a *loxP* Element at one end of the *aadA* Expression Cassette in 'pCP7-aadA1'

4.3.1.2. Introduction of a *loxP* Element into the *GFP* Open Reading Frame in pBAD-GFP

4.3.1.2.1. Construction of Derivatives of 'pBAD-lox-GFP' for Improved GFP Expression

4.3.1.3. Introduction of the ‘aadA-loxP’ Cassette into ‘pBAD-lox-GFP-insN’ to make the Final Construct

4.3.1.4. Use of ‘pBAD-lox-aadA-N-GFP’ as an Efficient Reporter for *in vitro* Cre Activity Assays

4.3.2. Development of Transgenic Tobacco Plants Expressing Cre Recombinase Targeted to the Chloroplast

4.3.2.1. Assembly of the *Agrobacterium* Transformation Vectors

4.3.2.1.1. Cloning of *Cre* and the *rbcS* Transit Peptide Sequences

4.3.2.1.2. Construction of pRok-Cre

4.3.2.1.3. Construction of pRokS-Cre

4.3.2.2. Analysis of Transgenic Tobacco Plants

4.3.2.2.1. PCR Analysis
4.3.2.2.2. Protein Analysis  
4.3.2.2.3. *In vitro* Cre-Activity Assay of Transgenic Plant Tissue.

4.4. Discussion  
4.4.1. Construction of ‘pBAD-lox-aadA-N-GFP’, the Target for assaying Cre Activity  
4.4.2. Transgenic Tobacco Plants Expressing Cre  
4.4.3. *In vitro* Cre Activity Assays  
4.4.4. Cre/loxP-mediated Excision in Chloroplasts  
4.4.5. Future Work
CHAPTER 1
GENERAL INTRODUCTION

1.1. The Chloroplast.

One of the more notable features that distinguishes plant from animal cells is the presence of plastids. Plastids can take on different physiological forms depending on the tissue type and developmental stage of the plant. These forms include the amyloplast, which stores starch; the chromoplast, which contains the carotenoid pigments found in flowers and fruits; the elaioplast, which stores oils, and the chloroplast, which contains the entire photosynthetic machinery in green tissues (Kirk and Tilney-Basset, 1978). Chloroplasts develop from small undifferentiated proplastids during leaf mesophyll cell differentiation.

1.1.1. The Chloroplast Genome.

The discovery of an independent DNA species in chloroplasts (Sager and Ishida, 1963) started what in recent years has become an accelerated interest in chloroplast molecular biology. Understanding the organisation and expression of chloroplast genomes has been aided by technical developments in gene cloning and DNA sequencing. The first physical map of chloroplast DNA was constructed for maize in 1976 (Bedbrook and Bogarad, 1976); the first chloroplast gene was cloned in 1977 (Bedbrook et al., 1977) and the first completely sequenced chloroplast genomes were published in 1986 for tobacco (*Nicotiana tabacum* L.; Shinozaki et al., 1986) and liverwort (*Marchantia polymorpha* L.; Ohyama et al., 1986). Recent years have seen the chloroplast genome sequences of a number of land plants and algae determined, including rice (*Oryza sativa* L.; Hiratsuka et al., 1989), pine (*Pinus thunbergii* Parl.; Wakasugi et al., 1994), maize (*Zea mays* L.; Maier et al., 1995), the green alga *Chlorella vulgaris* Beij. (Wakasugi et al., 1997) and lately those of *Arabidopsis thaliana* L. (Sato et al., 1999), evening primrose (*Oenothera elata* Knuth.; Hupfer, et al., 2000), the legume *Lotus japonicus* (Regal) K. Larsen (Kato et al., 2000), bread wheat (*Triticum aestivum* L.; Ogihara et al., 2002) and spinach (*Spinacia oleracea* L.; Schmitz-Linneweber et al., 2001). The complete genomic sequence of the unicellular cyanobacterium *Synechocystis* sp. PCC6803, a
primitive photosynthetic organism (Kaneko et al., 1996), and the minimal plastid genome from the non-photosynthetic parasitic *Epifagus virginiana* L. (Wolfe et al., 1992) have also been published. Such a wealth of both chloroplast and ancestral sequences has made it possible to make well-supported conclusions about the evolution of the plant kingdom.

### 1.1.2. Chloroplast Genome Structure.

Among land plants the chloroplast genome usually comprises a single, circular, double-stranded DNA molecule, 120-160 kilobase in length, divided structurally into a Large Single Copy (LSC) and Small Single Copy (SSC) region, separated from each other by inverted repeats (IR) present in two copies per genome. The inverted repeat accounts for most of the size variation in vascular plant chloroplast genomes. It can range in size from 6 to 76kb in length (Palmer, 1985). Some legume and conifer chloroplast DNAs are exceptions to this pattern and lack IRs. It is highly likely that the IR was present in the common ancestor of land plants and that one copy of the IR was lost in some legumes and conifers during evolution (Palmer, 1985).

The gene order of the tobacco chloroplast genome (Shinozaki et al., 1986) is generally accepted as being representative of most land plants, since very little variation between it and the plastid genomes of other higher plant species has been observed. This map has proved to be very useful for subsequent sequencing projects. However, an assumption that chloroplast genomes from all species will have the same gene order would not be advised. Chloroplast genome rearrangements have been demonstrated within the legume family (Palmer and Thompson, 1981) and a number of plastid gene transversions and duplications have been noted in rice (Hiratsuka et al., 1989). Also, new genes not present in vascular plant chloroplast genomes have been found in the plastomes of the Euglenophyta *Euglena gracilis* Klebs (Hallick et al., 1993) and the red alga *Porphyra purpurea* Agardh (Reith and Munholland, 1995).
1.1.3. Chloroplast Genes.

Vascular plant and algal chloroplast genomes contain approximately 120 genes, most of which encode components of the photosynthetic electron transport machinery and elements of the transcriptional and translational apparatus (i.e. the genetic system; Dyer, 1984).

1.1.3.1. Genes Encoding Polypeptides of the Photosynthetic System.

Higher plant plastomes contain 5 genes coding for components of Photosystem I, 14 for Photosystem II, 4 for the cytochrome b/f complex and 6 for the chloroplast ATPase (Sato et al., 1999). In addition, angiosperms have up to 11 plastid genes (ndhA-K) that encode a chloroplast localised NADH dehydrogenase. All of these ndh genes have been found to be absent or non-functional in the plastome of pine (Wakasugi et al., 1994). All chloroplast genomes except those of some parasitic plants contain a functional rbcL gene encoding the large subunit of ribulose bisphosphate carboxylase/oxygenase (RUBISCO), while the small subunit is encoded by a nuclear gene.

Additional photosynthetic genes have been found in the plastid genomes of non-angiosperm species: psaA is found in pine (Wakasugi et al., 1994), liverwort (Ohyama et al., 1986) and Euglena (Hallick et al., 1993); chlB, chlL and chlN are found in the gymnosperms (Wakasugi et al., 1994) and in numerous algae including Chlamydomonas reinhardii Dangeard (Li et al., 1993; Liu et al., 1993; Choquet et al., 1992; Suzuki and Bauer, 1992) and Porphyra purpurea (Reith and Munholland, 1995). The latter three genes encode a light-independent protochlorophyllide reductase activity, which allows chlorophyll production in the absence of light, a process that cannot occur in angiosperms (Suzuki and Bauer, 1995).

The plastome of higher plants also codes for a few polypeptides not directly involved in photosynthesis, such as acetyl-CoA carboxylase (from accD gene) and an ATP-dependent protease (encoded by clpP; Maurizi et al., 1990). They also contain a number of conserved open reading frames (ycfs), whose functions have yet to be determined.
1.1.3.2. Genes Encoding Polypeptides of the Genetic System.

All of plastid genomes sequenced to date (references cited above) contain the 4 genes that code for the “plastid encoded RNA polymerase” (rpoA, rpoB, rpoC1 and rpoC2). The higher plant chloroplasts contain at least 60 ribosomal proteins and approximately one-third of them are encoded by the plastid genome. In angiosperms, the plastid genome contributes 12 peptides to the 30S small subunit and 9 to the 50S large subunit of the ribosome (Shimada and Sugiura, 1991; Sato et al., 1999). Four plastid genes coding for ribosomal proteins contain introns in most plants with rps12 requiring trans-splicing during mRNA maturation. The rpl16 gene is missing from the plastomes of both pine and liverwort (Wakasugi et al., 1994; Ohama et al., 1986), demonstrating gene migration to the nucleus in plant lineages that emerged after the Bryophyta. Virtually all the ribosomal protein genes retained in the plastid genome have homologues in *Escherichia coli* that function in the early stages of the ribosome assembly pathway (Wolfe et al., 1992).

1.1.3.3. Genes coding for the tRNAs and rRNA.

The chloroplast also encodes a number of stable RNA species. Chloroplast ribosomes contain 23SrRNA, 16SrRNA, 5SrRNA and 4.5SrRNA (in land plants), all of which are plastid-encoded and their genes are generally clustered in the IR region (found in the SSC region of gymnosperms; Wakasugi et al., 1994). Gene order and sequence conservation among higher plants has been noted to be extremely high for these genes (much more so than for ribosomal protein genes), with a 98-100% sequence identity existing between *A. thaliana* and *N. tabacum* (Sato et al., 1999; see also Table 3.5. of this report).

Chloroplast translation involves a number of plastid-encoded transfer-RNA (tRNA) species. In *A. thaliana* a total of 30 tRNA genes (not including 7 duplicates in the IR region) were identified (Sato et al., 1999). The same number of tRNA genes was identified in the plastid genomes of tobacco (Shinozaki et al., 1986) and rice (Hiratsuka et al., 1989), but two extra genes were described in pine (*trnP(GGG)* and *trnR(CCG)*; Wakasugi et al., 1994) and one in liverwort (Ohyama
et al., 1986). Plastid tRNA sequences are highly conserved, with at least 80% homology existing between those genes of tobacco, rice and liverwort (Shimada and Sugiura, 1999). In land plants, the tRNA genes are scattered throughout the chloroplast genome. This differs from the situation in *Euglena gracilis* where most of the tRNA genes are clustered (Hallick et al., 1993). Four tRNA pseudogenes have been identified in rice. These are proposed to have resulted from multiple genome rearrangements, which must have occurred more than four times if the ancestral rice chloroplast genome was similar in gene order to that of tobacco (Shimada and Sugiura, 1989). Similar pseudogenes and genome rearrangements have also been reported in wheat (Howe, 1985). All 61 possible codons are used in chloroplast genes that encode polypeptides. Assuming that normal wobble base-pairing occurs in codon/anticodon recognition, the 30 plastid-encoded tRNAs are probably sufficient to read all codons (Sugiura, 1992).

1.2. Chloroplast Gene Expression.

1.2.1. Chloroplast Transcription.

The 120 or so plastid genes are organised into approximately 50 transcription units, suggesting that many chloroplast genes are transcribed polycistronically (Sugiura, 1992). Recent research has shown that transcription in the plastids of higher plants relies on two different transcriptional machineries, the plastid-encoded RNA polymerase (PEP) and the nuclear-encoded plastid RNA polymerase (NEP). In general NEP transcribes the *rpoB* operon (Liere and Maliga, 1999) and various housekeeping genes such as rRNAs (Sun et al., 1989), tRNAs and ribosomal protein genes (Hess et al., 1993), whereas PEP transcribes both the housekeeping and the photosynthetic genes (Mullet, 1993; Stern et al., 1997).

1.2.1.1. The Plastid-Encoded RNA Polymerase (PEP).

The PEP subunits α, β, β′ and β″ are coded for by the plastid genes *rpoA*, *rpoB*, *rpoC1* and *rpoC2*, which are homologous to genes encoding the eubacterial RNA polymerase (Pfannschmidt et al., 2000).
1.2.1.1. PEP Promoters.

A number of PEP promoters have been analysed in detail, including those of \textit{rbcL}, \textit{atpB} (Manen \textit{et al}., 1994), \textit{psbD} (Kim \textit{et al}., 1999) and 16S rRNA (Lerbs-Mache, 2000). Their sequences are highly similar to those of prokaryotic promoters, consisting of -35 and -10 consensus elements. The spacing between the two elements is typically between 17 and 19 base pairs and changes in this distance has been shown to affect promoter strength \textit{in vitro} (Link, 1994). Some PEP promoters have also been shown to contain sequences upstream of the core -35/-10 elements that are responsible for regulating transcription (Iratni \textit{et al}., 1994; 1997). The best studied of these is the blue-light-regulated promoter of the \textit{psbD-psbC} operon. A 22 bp sequence, termed the AAG-box, is understood to recruit the polymerase to the core promoter by binding to the AGF (AAG binding factor) protein complex (Christopher and Mullet, 1994; Kim and Mullet, 1995; Kim \textit{et al}., 1999; Allison and Maliga, 1995). AGF is therefore involved in transducing nuclear control over photosynthesis in response to changing light conditions.

1.2.1.1.2. Sigma (\(\sigma\)) Factors in Plastid Transcription.

Although the core subunits of the PEP are all plastid gene products, its transcriptional activity is still under the influence of the nucleus by means of nuclear-encoded sigma-like factors (SLFs) similar to the \(\sigma^{70}\) of \textit{E. coli} (reviewed in Allison \textit{et al}., 2000). They provide the PEP holoenzyme with promoter-selectivity and regulation in response to environmental and developmental stimuli. The SLFs in effect control the transcription of certain plastid genes by recognising sequence-specific elements in or adjacent to the promoter (Gross \textit{et al}., 1998). Isolation of 3 SLFs from mustard plastids during various developmental stages led to the discovery that all were more tightly bound to the PEP in etioplasts than in chloroplasts. This study also revealed that the same SLFs possessed greater affinity for the \textit{psbA} promoter compared to other plastid genes in etioplasts (Tiler and Link, 1993a). Further examination revealed that such functional differences were dependent on the phosphorylation state of the SLFs (Tiler and Link, 1993b). The discovery of a plastid transcription kinase, sensitive to redox potential, confirms the significance of SLFs in regulating photosynthesis (Baginsky \textit{et al}., 1999).
Database searches for conserved sigma factor domains have continued to identify new SLFs. So far higher plant screens have revealed 6 different SLF genes in Arabidopsis and 5 in maize, as well as single genes in rice, wheat, sorghum and mustard (Allison, 2000). Mixing the purified mustard SLF (encoded by SigI) with the E. coli core RNA polymerase provided direct proof of the function of the SLFs in promoter recognition. The purified SIG1 protein caused the E. coli holoenzyme to bind specifically to the plastid psbA promoter (Kestermann et al., 1998). It is clear that higher plants have numerous plastid-localised SLFs that are expressed in response to different environmental cues, similar to the bacterial model (Gross et al., 1998). The promoters of two Arabidopsis SLF genes, AtSigA and AtSigB, were used to drive expression of the GUS reporter gene in transgenic plants (Kanamaru et al., 1999). GUS expression assays revealed that both promoters were activated in green tissues such as leaves, cotyledons and hypocotyls but inactive in roots, seeds and flowers. AtSigB was also shown to be activated 24 hours earlier in cotyledons than AtSigA.

1.2.1.2. The Nuclear-Encoded Plastid RNA Polymerase (NEP).

In addition to the E. coli-like PEP, the existence of other biochemically distinct RNA polymerase activities in plastids has been known for some time (Howe, 1996). Deleting the rpoB gene encoding the essential β subunit of the PEP in tobacco, demonstrated the existence of a second, nuclear-encoded RNA polymerase activity in plastids (Allison et al., 1996). An examination of the resulting pigment-deficient and photosynthetically defective plants showed very low levels of mRNAs for the photosynthetic genes rbcL, psbA and psbD, due to the lack of PEP. In contrast, transcript levels for genes whose products comprise the genetic system (e.g. 16S rRNA and rpl16) accumulated to normal or even higher than normal levels. NEP is imported into the plastid (Weihe and Börner, 1999) and has biochemical properties similar to T7 phage RNA polymerase. It has been shown to recognise a site at or near the bacteriophage T7 promoter in vitro (Lerbs-Mache, 1993). The Arabidopsis nuclear gene RpoZ has been identified as the putative NEP catalytic subunit and was shown to be homologous to both the mitochondrial (RpoY) and phage RNA polymerases (Hedtke et al., 1997).
1.2.1.2.1. NEP Promoters.

Most plastid genes that are transcribed by NEP are also transcribed by PEP, although a few genes, such as accD and the rpoB operon, are transcribed exclusively by NEP (Hajdukiewicz et al., 1997; Silhavy and Maliga, 1998). All NEP promoters identified to date have been clearly shown to differ from the prokaryotic -10/-35 σ^70-type promoter transcribed by the PEP. Several motifs have been identified (Hubschmann and Börner, 1998; Silhavy and Maliga, 1998; Hajdukiewicz et al., 1997; Sriraman et al., 1998) and used to distinguish three classes of NEP promoter. NEP promoter Class Ia and Ib share a common region, referred to as the YRT-box immediately upstream of the transcription initiation site. This sequence consists of a moderately conserved AT-rich sequence adjacent to a highly conserved YRT core motif. Class Ib also has a conserved GAA motif, which is located approximately 10-20 nucleotides upstream of the YRT core of the YRT-box. In contrast Class Ia NEP promoters lack the GAA motif. The existence of another class of NEP promoters (Class II) that differ completely in sequence and organisation from the Class I promoters has also been demonstrated. The critical motif defining Class II NEP promoters was found to be present in liverworts and conifers, showing that NEP-dependent transcription appeared early in plant evolution (Sriraman et al., 1998a; reviewed by Weihe and Börner, 1999).

1.2.1.3. Transcription of the rRNA Operon.

Plastid rRNA transcription is one of the most studied examples of plastid gene expression. The tRNA^Val^-16SrDNA intergenic region of most higher plants contains four potential transcription start sites; two prokaryotic-type promoter elements, P1 and P2, and two non-consensus promoters PC and P3, shown to be transcribed by NEP (Lerbs-Mache, 2000). Primer extension analysis performed on the leaves of a number of closely related species (spinach and tobacco and Arabidopsis) revealed differing plastid rrn operon promoter (Prrn) usage, despite the sequence similarities (Sriraman et al., 1998b). Plastid rRNA transcription is initiated from P1 but not PC in tobacco; from PC but not P1 in spinach; and from both P1 and PC in Arabidopsis. These findings suggest that PC and P1 transcription activities are dependent on promoter-specific accessory factors, encoded in the nucleus and imported from the cytoplasm. Consequently, rrn
transcription can be highly regulated, allowing the nucleus to control the general rate of ribosome formation in the chloroplast. This conclusion is supported by another study that looked at tissue-specific PC activity in spinach (Iratni et al., 1997).

Recently Bligny et al. (2000) purified three different transcriptional activities from spinach chloroplasts: PEP, the conventional NEP (NEP-1), and a previously undescribed NEP-2. NEP-2 does not cross-react with NEP-1-specific antibodies and specifically transcribes the *rrn* operon from PC. Two different protein complexes have been shown to interact with the spinach plastid *rrn* promoter (Baeza et al., 1991; Iratni et al., 1994): a large complex of PEP in tight association with the transcription factor CDF2 and a small complex of NEP activity (NEP-2; Bligny et al., 2000) associated with CDF2. CDF2 provides dual regulation of both PEP and NEP *rrn* transcription. CDF2 acts as an initiation factor with NEP-2, but also represses transcription by recruiting inactive PEP to the P2 promoter (Bligny et al., 2000).

1.2.2. Chloroplast mRNA Processing.

Chloroplast gene transcripts proceed through several forms of processing before being translated. These steps include splicing, endonucleolytic cleavage, 5' and 3' end maturation and editing. Interest in understanding these processes is high because they, in conjunction with mRNA stability, are important points of regulation in plastid gene expression (reviewed Monde et al., 2000). Post-transcriptional regulation of plastid gene expression via mRNA stability has been studied intensively. Deng and Gruissem (1987) inferred an increase in psbA mRNA stability in leaves grown in the light compared to those grown in the dark.

Chloroplast transcription termination is highly inefficient, and typically requires both 5' and 3' end processing (Rott et al., 1996). Both the 5' and 3' untranslated regions (UTR) can affect mRNA stability. A number of chloroplast genes have been studied in detail in *Chlamydomonas*, namely petD (Drager et al., 1998), psbB (Vaistij et al., 2000) and psbD (Nickelsen et al., 1994), and in all cases the 5' UTR interacts with the nuclear genotype. Sequence-specific binding of
proteins to the 5' UTR was shown to either stabilise or destabilise these transcripts (Kuchka et al., 1989; Nickelsen et al., 1994).

The 3' UTR of almost all plastid mRNAs contains an inverted repeat (IR) that forms a stem loop structure (Belasco, 1993). The influence of the 3' UTR and IR on the transcript stability has been demonstrated by deletion studies carried out on the Chlamydomonas genes atpB (Stern et al., 1991) and psaB (Lee et al., 1996). Deletions of part or all of the 3' UTR sequence in transformed Chlamydomonas chloroplasts led to decreased atpB mRNA accumulation, whereas transcription rates were unaffected.

A sub-set of plastid mRNAs are also subjected to “editing” and the genomes of both maize (Maier et al., 1995) and black pine (Wakasugi et al., 1996) have been extensively investigated for the presence of mRNA editing sites. Both plastomes contained approximately 25 editing sites, corresponding to 0.02% of the entire genome. The vast majority of plastid mRNA editing events affect the second codon position, but they have also been found occasionally to create start codons (Hoch et al., 1991). Editing appears to be confined to pyrimidine bases, with C to U transitions being most commonly encountered (Maier et al., 1996). U to C reverse-editing occurs infrequently (Yoshinaga et al., 1996), but no other type of nucleotide substitution has been detected in plastid transcripts. The editing mechanism is most probably similar to that which occurs in mitochondria (Yu and Schuster, 1995), where the base change is the result of enzymatic conversion rather than nucleotide excision and replacement. Transcript sequences flanking the editing site are directly involved in recognition by the editing machinery (Boch et al., 1996). Surprisingly, organellar RNA editing does not appear to be a processing step of ancient origins. While it does occur in all higher plant lineages, it is notably absent in all algae and some Bryophytes (Freyer et al., 1997).

1.2.3. Chloroplast Translation.

Because of the prokaryotic origins of plastids it is no surprise that “Shine-Dalgarno” (SD) sequences are considered to play some part in plastid translation initiation (Dron et al., 1982; Maidak et al., 1996). In bacteria a SD sequence is
typically found at about ten nucleotides upstream (-10) relative to the translation initiation codon (McCarthy and Brimacombe, 1994). Analysis of vascular plant sequences has revealed similar elements in the 5' untranslated region of many chloroplast genes (Bonham-Smith and Bourque, 1989). *In vitro* toe-print assays established that plastid 30S ribosomes associate with a SD-like sequence in barley *rbcl* and *psbA* mRNAs (Kim and Mullet, 1994). However, not all plastid genes contain a recognisable SD-like sequence, suggesting that its presence in the 5' UTR is not an absolute requirement for plastid translation initiation (Fargo et al., 1998). Mutation analysis of the SD-like elements of tobacco *psbA* (Hirose and Sugiura, 1996) and *rps14* (Hirose et al., 1998) has clearly demonstrated the variable importance of the consensus SD element in plastid genes: an intact SD-like sequence is required for translation in *rps14* but not *psbA*. *Chlamydomonas* plastid genes generally lack a true SD-like consensus element, instead exhibiting an A–U rich (85%) consensus in the 5' UTR (Koo & Spremulli, 1994). This base composition may reduce RNA secondary structure and facilitate ribosome access to the translation initiation codon (Koo and Spremulli, 1994; Betts and Spremulli, 1994).

Recent research indicates that SD-like sequences function in conjunction with other *cis*-acting elements that interact with nuclear-encoded factors (Barkan, 1993; Hauser et al., 1996). The 5' UTR of the *Chlamydomonas petD* mRNA is understood to contain three distinct sequence elements that influence its translation (Sakamoto et al., 1994; Higgs et al., 1999). Similar findings were obtained following mutational analysis of *psbC* (Zerges et al., 1997) and *rps7* (Fargo et al., 1999). The tobacco *psbA* 5' UTR contains three elements, all located within 30 nucleotides upstream of the start codon, required for correct translation initiation (Hirose and Sugiura, 1996). Although this leader does not contain a SD-like consensus, two of the elements (RBS1 and RBS2) are believed to function in a similar fashion by binding to the 3' end of 16S rRNA. The third element, an AU-rich box, binds several protein factors encoded by nuclear genes (Hirose and Sugiura, 1996).
A number of nuclear-encoded proteins have been identified that participate in plastid gene translation. Numerous examples of these have been discovered in *Chlamydomonas* (Zerges, 2000). In maize the nuclear-encoded *atpI* gene is required for the translation of the plastid *atpB* operon (McCormac and Barkan, 1999) and maize *crpI* mutants have been shown to disrupt *petA* and *petD* expression at the level of translation (Fisk *et al.*, 1999). Examples of translational regulation in plastids in response to both the developmental state of the plant (Zhao *et al.*, 1999) and changing light conditions (Bruick and Mayfield, 1999) have also been reported.

1.3. Chloroplast Genome Evolution.

1.3.1. Chloroplast Origins.

Chloroplasts, like mitochondria are believed to have originated from an endosymbiotic event between a prokaryotic-like organism and an ancient precursor of the eukaryotes approximately 2,000 million years ago (Margulis, 1970). This theory was supported at the molecular level by the early observations that the nucleotide sequences of plant chloroplast 5S rRNA and 16S rRNA genes were highly similar to those of the cyanobacterium *Anacystis nidulans* (Dyer and Bowman, 1979; Tomioka and Sugiura, 1983), and later by comparisons between the completely sequenced genome of the blue-green alga *Synechocystis* and various higher plant plastid genomes (Kaneko *et al.*, 1996; Martin *et al.*, 1998).

Ultrastructural studies have demonstrated the existence of plastid organelles surrounded by two, three and even four membranes (Gibbs, 1978; Gibbs, 1981). Those surrounded by two membranes are thought to have arisen by primary endosymbiosis, whereas those surrounded by three and four are thought to have arisen by a process known as secondary endosymbiosis, whereby a eukaryotic organism (most likely unicellular) engulfed a second photosynthetic eukaryote (Gibbs, 1978). Combined plastid and nuclear nucleotide sequence analyses have also supported this theory (Bhattacharya and Medlin, 1998).
Plastid genome sequence analysis is proving to be an exceptionally powerful tool for the determination of phylogenetic relationships between various photosynthetic organisms. Analysis of plastid rRNA gene sequences from a wide variety of species has provided convincing evidence concerning evolutionary theories (Bhattacharya and Medlin, 1998). The green algae and land plants form a monophyletic lineage (the chlorophytes). The chloroplasts of the chlorarachniophytes (four membranes) and the euglenophytes (three membranes) can be traced to a secondary endosymbiotic event involving an ancestor of the chlorophyte lineage. A phylogenetic tree of the chlorophyte lineage, including a number of higher plants is shown in Figure 1.1. Similarly it has been postulated that the haptophytes and heterokonts originated via secondary endosymbiosis of an ancient member of another lineage, that of the red algae (rhodophytes).

Recent reports based on sequence analysis have concluded that it is highly likely that the divergence of all primary plastids (chlorophytes, rhodophytes and glaucocystophytes) can be traced back to a unique endosymbiotic event (Kowallik, 1994; van de Peer et al., 1996; Turner, 1997; Bhattacharya and Schmidt, 1997). The presence of common gene clusters, not found in cyanobacteria, indicates that they were grouped together subsequent to endosymbiosis (Reith, 1995). Molecular data has also been used to support the theory that all members of the chlorophytes have evolved from a single unicellular green alga to more complex algae, to the bryophytes and then to the higher land plants (Bhattacharya and Medlin, 1998).

Methodological problems in algorithms and conflicting results are realities that must be accepted when phylogenetic studies are based on single genes. However, with the completion of numerous plastid genome sequences, Martin et al. (1998) were able to draw detailed phylogenetic conclusions based on a wider range of genetic markers. A total of 46 protein-coding plastid genes, common to the ten species were selected and used to construct phylogenetic trees (see Figure 1.2). Similarly, Lemieux et al. (2000; see Figure 1.3), having sequenced the chloroplast genome of the flagellate Mesostigma viride Lauterb., suggested it belonged to the earliest diverging green plant lineage discovered to date. Mesostigma represents a lineage that emerged before the divergence of the Streptophyta (land plants and
Figure 1.1. Small Subunit rDNA Phylogeny of the Chlorophyte Lineage. This tree was constructed by Bhattacharya & Medlin (1998), with a weighted maximum parsimony method, and the results of bootstrap analysis (200 replications) are shown as branch nodes of differing thickness (see box on the right). The likely position of divergence of the prasinophyte *M. viride* is shown with a broken line. The phylogeny is rooted within the branch leading to the rDNA sequence of the glycocystophyte *C. paradoxa*. 
Figure 1.2. Plastid Phylogeny Interpreted from Chloroplast Proteins, based on Martin et al. (1998). Complete chloroplast genome sequences of *Euglena gracilis* (X70810; eug), *Porphyra purpurea* (U38804; por), *Odentella sinensis* (Z67753; odo), *Cyanophora paradoxa* (U30821; cya), *Zea mays* (X86563; zea), *Oryza sativa* (X15901; ory), *Nicotiana tabacum* (S54304; nic), *Pinus thunbergii* (D17519; pin), *Marchantia polymorpha* (X04465; mar) as well as *Synechocystis* sp. Strain PCC6803 were obtained from GenBank. It must be pointed out that this figure is only approximately demonstrative of evolutionary distances. A more detailed tree, calculated using Dayoff distances may be found in Martin, et al. (1998).

Figure 1.3. Phylogenetic Position of *Mesostigma* as infered from 53 Chloroplast Proteins. This tree was constructed by Lemieux et al. (2000) with distance, maximum parsimony and maximum likelihood inference methods using the glaucocystophyte *Cyanophora paradoxa* proteins as an outgroup.
their closest green algal relatives; Bremmer, 1985) and Chlorophyta (the rest of green algae; Sluiman, 1985). This study also demonstrates that the structure and gene organisation of the chloroplast genome has been extremely well conserved in the ancestral line leading to land plants. In order to correctly interpret these plastid phylogenetic studies it must be noted that chloroplast genomes appear to have evolved considerably faster than their cyanobacterial homologues (Palmer and Delwiche, 1998).

1.3.2. Gene Transfer to the Nucleus.

During the course of its development from a free-living cyanobacterial ancestor into the highly specialised chloroplast, the organelle has lost much of its genetic independence. Compared to the 3,000 genes found in the blue-green alga *Synechocystis* sp. strain PCC6803 genome (Kaneko et al., 1996) only 100-200 are found in plastid genomes. Over the course of plant evolution most of the ancient chloroplast genes have been lost or translocated to the nucleus (Oliver et al., 1989; Baldauf and Palmer, 1990). Ancestral plastid genes that came under the control of the nuclear/eukaryotic expression machinery acquired a transit peptide sequence which facilitated import of the encoded proteins from the cytoplasm into the chloroplast. The transit peptide portion of these proteins is generally contained at the N-terminal end, within the first 100 amino acids. Although transit peptide sequences are poorly conserved among various chloroplast-targeted proteins (indicating exclusive gene-evolution), they do have certain features in common, such as numerous hydroxylated amino-acid residues and few negatively charged residues (Keegstra and Cline, 1999). Numerous computer programs have been developed that can predict organellar transit peptide sequences, including ChloroP (Emanuelson et al., 1999), TargetP (Emanuelson et al., 2000) and Predotar (unpublished data; see http://www.inra.fr/servlets/WebPredotar). Membrane lipids found in the outer-envelope of the chloroplast are known to interact with transit peptide sequences (Pinnaduwage and Bruce, 1996). Both the translocation mechanism and the protein-import machinery resemble transport systems found in cyanobacteria (Heins et al., 1998).
Shih et al. (1986) presented strong evidence at the molecular level showing that genes coding for chloroplast enzymes could have been transferred from the genome of a prokaryotic symbiont to the nucleus. This study demonstrated that the chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH), both nuclear encoded, evolved from different lineages that diverged approximately 1,700 million years ago. Considering that the divergence between animal and plant genes coding for cytosolic enzymes occurred later, at about 850 million years ago it provides clear evidence in support of the Symbiotic Theory.

Martin et al. (1998) recorded the presence or absence of 210 protein-coding genes in nine completely sequenced chloroplast genomes as well as their homologues from the Synechocystis sp. PCC6803 genome (Kaneko et al., 1996) and found clear evidence of 44 bona fide cases of the transfer of plastid genes to the nucleus. The majority of the genes relocated to the nuclear genome were found to have undergone parallel loss in independent lineages and these outnumbered phylogenetically unique losses by more than 4:1. Although most genes were transferred from the ancestral plastome to the nucleus early in plant evolution, a number of gene transfers occurred relatively recently in various angiosperm lineages. Loss of chloroplast rpl22 in legumes was due to a recent transfer of genetic material to the nucleus (Gantt et al., 1991) whereas accD loss from the plastome of grasses is a case of gene substitution (Konishi et al., 1996). Numerous independent transfer events led to the transfer of plastid infA to the nucleus during angiosperm evolution (Millen et al., 2001).

A detailed homology-based analysis of the non-redundant Arabidopsis Database was undertaken in order to estimate how many nuclear genes are derived from cyanobacteria (Abdallah et al., 2000). Utilising the ChloroP program to detect transit peptide-like sequences, Emanuelson et al. (1999) predicted that between 1900 and 2500 (12%) of all nuclear-encoded proteins are targetted to the chloroplast. This number does not include those proteins encoded by the plastid genome or chloroplast proteins that don’t have a recognisable transit peptide. Approximately 35% of the chloroplast-targeted proteins showed significant
homology to proteins encoded by the *Synechocystis* genome. It is most likely that the endosymbiont's contribution was actually much greater than this estimate, but plant-specific genes that have evolved since the origin of the plant kingdom also participate in chloroplast function.

The question as to why some cyanobacterial-derived genes were transferred to the nucleus, whereas others were retained in the plastome, can only be answered by reasonable postulation. It could be that it is easier for certain hydrophobic proteins to be directly inserted into membranes (e.g. thylakoid membranes) following translation within the chloroplast as opposed to importing them from the cytoplasm. Analysis of both engineered and natural proteins expressed in the nucleus and targeted to the inner membrane of the mitochondria revealed the importance of hydrophobicity in the import process (Claros et al., 1995). Also the expression of some genes within the chloroplast allows for a direct and rapid means of regulation in response to redox changes (Pfannschmidt et al., 1999).


The last decade has witnessed the development of various procedures for transformation of the chloroplast genome of a number of plant species. The first report of a successful chloroplast transformation was in the unicellular alga *Chlamydomonas reinhardtii* (Boynton et al., 1988) while the first higher plant to give rise to stable plastid transformants was tobacco (Svab et al., 1990). Plastid transformation itself is a multistep process, in which a selectable marker gene flanked by homologous chloroplast targeting sequences is introduced into the plastid genome. Transformant plastids are selected during regeneration of the plant tissue and screened using standard genetic test procedures. Each step in the process along with its underlying principles is outlined below.

1.4.1. DNA Delivery into Chloroplasts.

As of yet an *Agrobacterium*-based system for foreign DNA delivery and insertion into the plastid genome is not available as is the case for nuclear transformations. A successful chloroplast delivery system should allow passage of
foreign DNA not only through the cell wall and plasma membrane but also through the double membrane of the plastid envelope, with minimal sub-cellular damage so that transformed cells can continue to replicate. Currently there are a number of such systems available.

1.4.1.1. Biolistic Bombardment.

Although developed for nuclear transformation (Klein et al., 1988), the biolistic bombardment of plant tissue with microprojectiles (tungsten or gold particles) coated with DNA is currently the most widely used method of DNA delivery to the plastid. This technique has been utilised to obtain stable transformants for a variety of species including tobacco (Svab et al., 1990), tomato (Ruf et al., 2001), potato (Siderov et al., 1999), Arabidopsis (Sikdar et al., 1998) and rice (Khan and Maliga, 1999) as well as transient transformants for carrot, marigold and red pepper (Hibberd et al., 1998). The major attraction for choosing this system is for its technical simplicity; entire organs such as leaves can be bombarded and subsequently regenerated on a selective medium. Unfortunately, as well as being a comparatively expensive technique, the non-targeted delivery of the DNA into cells means there is an unavoidable risk of integration into the nuclear genome.

1.4.1.2. Polyethylene Glycol Treatment.

Polyethylene glycol (PEG) treatment of protoplasts has been shown to successfully deliver DNA to the chloroplast resulting in stable plastid transformation (Golds et al., 1993; Koop et al., 1996). This technique is a less expensive alternative to the biolistic method and with as many as 20 to 40 transformants per $10^6$ treated protoplasts (Koop et al., 1996), its efficiency is impressive. However, obtaining and manipulating protoplasts requires specialised tissue culture skills and to date successful transformation protocols have been demonstrated only for tobacco (Golds et al., 1993) and its close relative Nicotiana plumbaginifolia Viv. (O’ Neill et al., 1993).
1.4.1.3. The Femtoinjection Technique.

Recent reports have presented a potentially exciting alternative DNA delivery system that may overcome the problem of nuclear insertion and many other problems associated with the currently used methods (Knoblauch et al., 1999). This technique uses a sub-micron diameter glass syringe. The DNA-containing fluid in the tip is expelled into the plastid via the thermal expansion of a liquid metal alloy called galinstan. Using this technique, the gene for green fluorescence protein (GFP) under the control of the chloroplast ribosomal RNA promoter (Prn) was injected into plastids and subsequent transient expression demonstrated. Importantly the cells survived any damage caused by the manipulation. The authors of this publication observed an interesting phenomenon of movement of green fluorescence to neighbouring, uninjected chloroplasts. Although in the early stages of investigation, it is thought that this is due to the formation of transient envelope-derived tubules called stromules between plastids. These may play a role in the establishment of homoplasmy following the initial plastid transformation event.

1.4.1.4. Future Possibilities in DNA Delivery to the Plastid.

Because plant transformation limitations are mainly due to tissue regeneration difficulties, many laboratories are pursuing lines of research that are looking at DNA delivery to tissues that will ultimately give rise to gametes (Chee and Slighton, 1995; Birch, 1997), thus eliminating the tissue culture requirement. Biolistics have been used to directly transform apical meristem cells that subsequently grow and ultimately produce seeds (Birch, 1997). Microinjection of DNA into ovaries leading to the production of transformed embryos has also been reported (Zhou et al., 1983; Hu and Wang, 1999). Although these techniques have been directed at nuclear transformation, it may be possible to modify them for application to chloroplast transformation.

1.4.2. Integration of Foreign DNA into the Plastid Genome.

Integration of foreign DNA into the plastid genome occurs via homologous recombination between the flanking sequences of the transformation vector and the plastid genome. A homolog of the E. coli RecA protein, known to promote
homologous recombination in bacteria, has been detected in the plastids of higher plants (Cerutti et al., 1992). Assuming that only one or at most a tiny proportion of the many plastid DNA molecules in the plant cell have successfully integrated the transforming DNA, it is essential to put in place a system that allows selection of the transformant over wild type copies of the genome. Such selection should ultimately lead to plant tissue that is 'homoplasmic' for the transformed genome (i.e. all copies identical), a requirement for the long-term stability of the transformed plastid line and the interpretation of research results.

1.4.3. Selectable Markers.

Early work on *Chlamydomonas* plastid transformation utilised photosynthetic mutant strains with altered, non-functional *atpB* or *psbA* genes as the recipients for cloned DNA containing the wild type, functional copy of the gene. Transformant selection was based on the restoration of photosynthetic growth (Boynton et al., 1988; Boynton et al., 1990). Subsequent selection procedures were based on the introduction of antibiotic resistant genes (e.g. the *aadA* gene) into the *Chlamydomonas* chloroplast genome (Goldschmidt-Clermont, 1991). Such antibiotic selection systems also proved to be applicable to higher plant chloroplast transformations (Svab et al., 1990; Svab and Maliga, 1993).

1.4.3.1. Chloroplast Mutations conferring Resistance to Antibiotics.

The first plastid transformation experiments in higher plants utilised mutant alleles of the chloroplast 16S rRNA gene to confer resistance to the antibiotics spectinomycin and streptomycin (Svab and Maliga, 1991). Since then numerous other mutations found on the 16S rRNA, the 23S rRNA and *rps12* genes have been identified as potential marker gene candidates, conferring resistance to lincomycin, spectinomycin and streptomycin, respectively (Kavanagh et al., 1994). The utilisation of two linked, independent mutation markers makes it easier to distinguish successful plastid transformants from spontaneous mutational 'escapes'. Both markers can be located within the same gene (Svab et al., 1990) or in two separate but adjacent genes (Kavanagh et al., 1999).
Such antibiotic resistance markers are said to confer 'binding type' resistance to antibiotics because the molecular structure of the ribosome is altered in such a way, that it no longer binds the antibiotic and thus maintains its translational capabilities. The recessive nature of these mutant alleles explains the low transformation frequencies observed (Svab et al., 1990). However, the few resistant clones that are obtained reach homoplasmy quickly. This is probably due to the fact that resistant plastids do not inactivate the antibiotic and thus provide no cross-protection to susceptible plastids in cells containing a mixed population, as may be the case for a dominant selectable marker coding for an antibiotic detoxifying gene.

1.4.3.2. Heterologous Gene Markers

The *aadA* gene, encoding amino-glycoside 3' adenyltransferase, causes spectinomycin and streptomycin resistance in bacteria by modifying the structure of the antibiotic (Chinault et al., 1986). Expression of a chimeric *aadA* gene from the plastid genome has been shown to confer effective resistance to both these antibiotics in *Chlamydomonas* (Goldschmidt-Clermont, 1991), tobacco (Svab and Maliga, 1993) and other higher plant chloroplasts (Siderov et al., 1999; Sikdar et al., 1998; Khan and Maliga, 1999). The use of such a dominant selectable marker gene resulted in higher frequencies of recovered plastid transformants, although multiple rounds of shoot regeneration were required in order to obtain homoplasmic plants (Svab and Maliga, 1993).

Other chimeric antibiotic resistance genes have also been utilised for plastid transformation. Tobacco shoots resistant to high levels of kanamycin (500μg/ml) were obtained by introducing the neomycin phosphtransferase (*NPT II*) gene, under the control of plastid expression signals into the chloroplast genome (Carrer et al., 1993). Similarly, another bacterial gene *aphA-6*, which codes for an aminoglycoside phosphotransferase, was successfully used to select for *Chlamydomonas* plastid transformants on kanamycin and amikacin (Bateman and Purton, 2000). As yet, its use has not yet been extended to higher plant plastid transformation.
A negative selectable marker system, utilising the bacterial codA gene encoding cytosine deaminase has been shown to work for tobacco chloroplast transformation selection (Serino and Maliga, 1997). If expressed in the plastid, the enzyme will convert 5-fluorocytosine (added to the growth medium) into the toxic compound 5-fluorouracil, an inhibitor of pyrimidine biosynthesis. Shoots that thrive in the absence of 5-fluorocytosine but die in its presence are considered candidate transformants.

The increasing concern about the widespread use of antibiotics in genetically modified crops has prompted researchers to look for more environmentally friendly marker genes. The betaine aldehyde dehydrogenase (BADH) gene from spinach has recently been shown to be a more than adequate substitute for the bacterial antibiotic resistance marker (Daniell et al., 2001). The selection process involves conversion of toxic betaine aldehyde (BA) by the plastid expressed BADH enzyme to non-toxic glycine betaine, which also acts as an osmoprotectant (Rathinasabapathy et al., 1994). BA resistant clones regenerated within 12 days, much quicker than for spectinomycin resistance, which typically forms green shoots in about 45 days. A 25-fold higher transformation efficiency was also demonstrated with as many as 23 resistant shoots observed from a single bombarded leaf disk (Daniell et al., 2001). This landmark development could be a significant step closer to the implementation of chloroplast transformation technologies in economically important crops, including those that are naturally resistant to spectinomycin (e.g. maize).

Chloroplast-localised expression and detection of the jellyfish green fluorescent protein (GFP) presents the plastid engineer with the possibility of selecting transformed cells on the basis of their colour (Khan and Maliga, 1999; Sidorov et al., 1999). Unfortunately, the problem of overcoming heteroplasmy means its use alone in a transformation system is currently limited. A possible solution to this could be to temporarily reduce the number of chloroplasts per cell and the number of plastome copies per organelle prior to transformation. The number of chloroplasts per cell was shown to drop from 100 to one when antisense DNA for a gene controlling plastid division was used in A. thaliana (Osteryoung et
Additionally, when *Chlamydomonas* was grown in the presence of the thymidylate synthase inhibitor fluorodeoxyuridine the levels of chloroplast DNA were reduced to one-seventh the normal value (Wurtz *et al.*, 1977).

### 1.4.3.3. Marker Gene Removal from the Chloroplast.

Much of the public's fear about the use of antibiotic resistance genes in plant transformation strategies stem from the hypothesis that they could inactivate oral doses of the clinically administered antibiotics or that the genes themselves could be transferred to pathogenic microorganisms in the gastrointestinal tract or in the soil. Studies have been carried out to test the viability of such concerns and were unable to detect any incident where horizontal gene flow might have occurred (Syvanen, 1999). These findings do not, however, take from the fact that their persistence in the field would be unacceptable in such a consumer-driven market and that following selection of transformants they are clearly unnecessary, serving no function in the final engineered organism. In the year 2000 the German government banned the release of genetically modified (GM) crops containing antibiotic resistance genes (Peerenboom, 2000) and it is widely expected other EU states will follow suit.

Chloroplast transgenics is also responding to these concerns. The high plastid genome copy number and prokaryotic expression signals used may increase the probability of marker gene transfer to bacteria. To date, dominant antibiotic resistance genes have been the selection system of choice but geneticists are investigating ways of removing the marker genes following regeneration of transformant lines, in addition to developing novel transformation systems free of antibiotic selection (discussed above). Marker gene removal systems, be they for antibiotic resistance genes or otherwise, have the added advantage of facilitating marker gene re-use for a subsequent transformation event. This could be very useful because selectable markers are in limited supply.

By monitoring the occurrence of homologous recombination between direct repeats flanking a marker gene cassette, loss of the cassette has been demonstrated following removal of the selective growth conditions. This has been observed in
both *Chlamydomonas* (Fischer *et al.*, 1996) and tobacco plastid transformants (Iamtham and Day, 2000). For the tobacco experiment a 4.9kb fragment containing three genes, *aadA*, *bar* (confers resistance to the herbicide glufosinate) and *uidA* (codes for β-glucuronidase), was integrated into the plastid genome and homoplasmic plants selected. Removal of various gene combinations was demonstrated, depending on the selection medium applied and the positioning of the flanking repeats. The use of glufosinate alone in the selection medium was sufficient to obtain T₀ clones homoplasmic for an *aadA-uidA* free plastome. Even more interesting was the production of homoplasmic T₂ seedlings free of both *aadA* and *bar* in the absence of any selection (Iamtham and Day, 2000). Various repeat lengths have been tested for intervening sequence removal efficiency and shown to be a critical factor (Fischer *et al.*, 1996).

Generation of marker gene-free transplastomic *Chlamydomonas* has also been achieved by co-transformation of the marker gene and the transgene of interest on separate plasmids (Fischer *et al.*, 1996). This strategy utilises the fact that certain plastid genes are essential for plant cell survival. Interruption of such a gene with the *aadA* marker plus the presence of spectinomycin in the regeneration medium will only ever result in a genome heteroplasmic for both mutant and wild type copies. A second plasmid, containing the desired transgene could be transformed simultaneously and homoplasmic mutants isolated (homoplasmic for the transgene; heteroplasmic for the marker gene). Loss of the *aadA* cassette followed relief of the selective pressure with the transgene persisting in a homoplasmic state. Co-transformation techniques have also been demonstrated in higher plant chloroplasts (Carrer and Maliga, 1995).

The co-transformation or “essential ORF” method of marker gene recycling produces a clean deletion, without any non-functional foreign sequences remaining whereas with the “direct repeat” technique, a single copy of the flanking repeat remains in the plastome. On the other hand the direct repeat method is much more straightforward and does not require the same degree of screening for the presence of the transgene (Fischer, *et al.*, 1996). The use of site-specific recombinases to eliminate marker genes following nuclear transformation was achieved a decade
ago (Dale and Ow, 1991; Russel et al., 1992). Very recently a Cre recombinase system for chloroplast transformation marker gene removal was described (Hajdukiewicz et al., 2001; Corneille et al., 2001). This along with similar investigations performed during my Ph.D programme are described in Chapter 4.

1.4.4. Investigating Chloroplast Metabolism via Transformation.

Targeted interruptions of chloroplast genes via the transformation techniques discussed above have also provided scientists with a means of understanding gene function in vivo (Rochaix, 1997). Such “reverse genetics” studies performed on *Chlamydomonas* have proven successful for studying non-essential chloroplast genes. For example, the function of the plastid gene *psaC* encoding the iron sulphur protein of photosystem I was deduced from biochemical analysis of *Chlamydomonas* transformants, whose *psaC* coding sequence was disrupted by an *aadA* gene cassette (Takahashi et al., 1991). Similar gene knockout transformations have facilitated the elucidation of the function of many other *Chlamydomonas* chloroplast genes, including *petL* (*ycf7*; Takahashi et al., 1996), *cemA* (*ycf10*; Rolland et al., 1997), *ycf3* and *ycf4* (Boudreau et al., 1997). Conclusions drawn from these studies meant that similar functions could be attributed to homologous coding sequences in higher plant plastids (Rochaix, 1997). In addition, gene disruptions performed on the tobacco plastome provided definitive evidence for the existence of the nuclear-encoded plastid-targeted RNA polymerase (NEP) transcription apparatus (Allison et al., 1996) and for determining the function of the chloroplast NDH complex. Inactivation of tobacco *ndhB*, *ndhC*, *ndhK* and *ndhJ* showed that they were not essential for plastid function (Borrows et al., 1998; Kofer et al., 1998; Shikanai et al., 1998; Horvath et al., 2000). However, analysis performed on such mutant plants led to the hypothesis that the NDH complex is involved in cyclic electron flow around Photosystem I and thereby contributes to extra-ATP production, under adverse (e.g. photorespiratory) conditions (Horvath et al., 2000; Joët et al., 2001).

Site-directed mutagenesis of the plastid-encoded *rbcL* gene in tobacco has been used to alter the catalytic activity of the Rubisco holoenzyme (Whitney et al., 1999). Nucleotide substitutions were obtained by homologous recombination with
a linked \textit{aadA} gene, inserted downstream of the altered \textit{rbcL} coding sequence, giving the mutant a selective advantage in medium containing spectinomycin. Rubisco activities were compared between mutant and control transformants. Such targeted manipulations of plastid genes could be used to alter the metabolism of the plant in ways that could prove useful for agriculture. The engineered mutant \textit{rbcL} plants generated by Whitney \textit{et al.} (1999) were observed to be better suited to higher CO$_2$ conditions than their wild type precursors.

\textbf{1.5. Biotechnological Applications of Plastid Transformation.}

Since the first heterologous gene was successfully transformed into and expressed from the tobacco plastome (Svab and Maliga, 1993), numerous others have followed. Transplastomic technologies have advanced sufficiently in recent years that the chloroplast can now be viewed as both a viable target for agronomically advantageous traits and as a compartmentalised factory for chimeric protein production.

\textbf{1.5.1. Pest Resistance.}

Insect resistant plants have been engineered by expression of proteins that are toxic to the target insect when ingested. The Crystal toxin proteins (\textit{Cry}) from \textit{Bacillus thuringiensis} (\textit{Bt}) are potent insecticides active against most Dipteran, Lepidopteran and Coleopteran species (Hofte and Whiteley, 1989). Enormous difficulties were initially encountered in achieving a sufficiently high level of \textit{Bt} toxin expression from nuclear transgenic plants (less than 0.8\% of total soluble leaf protein; Wong \textit{et al.}, 1992). Factors such as aberrant translation, transcript instability (Murray \textit{et al.}, 1991) and inefficient codon usage (Perlak \textit{et al.}, 1991) were only overcome by constructing synthetic versions of the coding sequence that would be better suited to the nuclear expression environment in plants as opposed to the original prokaryotic source (Adang \textit{et al.}, 1993).

Because of its prokaryotic-like transcription/translation apparatus, high-level expression of an unmodified \textit{Cry1A} gene in the chloroplasts of tobacco proved to be much more straightforward (McBride \textit{et al.}, 1995). Indeed, these \textit{Bt
transplastomic plants accumulated the toxin to between 3 and 5% of total soluble leaf protein. The leaves of these plants were extremely toxic to the larvae of *Spodoptera exigua*, *Helicoverpa zea* and *Heliothis virescens* (McBride *et al.*, 1995). Even more impressive were the results obtained following the stable insertion of the *Cry2Aa2* operon into the plastid genome of tobacco (De Cosa *et al.*, 2001). The foreign protein accumulated to 46.1% of the total soluble leaf protein and remained stable in mature and older leaves. The marked improvement in *Bt* toxin expression and persistence is thought to be due to the presence of a gene (ORF2; located immediately upstream of *Cry2Aa2* gene in the three-gene operon) that encodes a *Cry*-specific chaperon protein, which resulted in the formation of folded protoxin crystals in chloroplasts. Despite the hyperexpression, *Cry2Aa2* transplastomic plants are in no way morphologically different from wild type (De Cosa *et al.*, 2001).

1.5.2. Herbicide Tolerance.

The herbicide glyphosate has been effectively used to control the majority of monocot and dicot weeds. Its biochemical mechanism is based on the competitive inhibition of 5-enol-pyruvyl shikimate-3-phosphate synthase (EPSPS), an enzyme involved in the aromatic amino acid biosynthetic pathway (a process that is compartmentalised within chloroplasts). Glyphosate resistance has been conferred on otherwise sensitive plants using nuclear transformation, both by overproduction of EPSPS (Shah *et al.*, 1986) and expression of a mutant version of the enzyme that does not bind the herbicide (della-Cioppa *et al.*, 1987).

The targeted insertion of the petunia EPSPS coding sequence into the plastid genome of tobacco resulted in plants that survived glyphosate concentrations ten times higher than the wild type lethal dose (Daniell *et al.*, 1998). Unfortunately, in spite of their increased herbicide tolerance, the tranplastomic plants were still only tolerant to a maximum of 8 oz acre\(^{-1}\) of the commercial herbicide Roundup\(^{\circledast}\), a figure well below the 64 oz acre\(^{-1}\) recommended for weed control. The nuclear expressed petunia enzyme confers low-level glyphosate tolerance and its codon usage is not suited for expression in plastids. Therefore the idea that high-level herbicide tolerance could be achieved using a bacterial gene
was investigated. Recent findings have clearly demonstrated that transplastomic lines expressing the EPSPS gene from *Agrobacterium*, *Achromobacter* and *Bacillus* species all show elevated glyphosate resistance (Ye *et al.*, 2001), as high as 128 oz acre\(^{-1}\) in the case of the gene from *Agrobacterium* strain CP4. Increased tolerance has been attributed to increased expression of heterologous EPSPS in chloroplasts (up to 10% total soluble protein). Interestingly Ye *et al.* (2001) also showed that codon optimisation was not critical to the large increase in protein expression, which was due instead to modification of translational control signals.

Another bacterial herbicide resistance gene that has been expressed in the chloroplast is the *bar* gene (Iamtham and Day, 2000; Lutz *et al.*, 2001). This gene from *Streptomyces hygroscopicus* codes for a phosphinothricin acetyltransferase (PAT), an enzyme that detoxifies phosphinothricin (PPT), the active component of the non-selective herbicide bialaphos (Thompson *et al.*, 1987). PPT acts as an inhibitor of glutamine synthetase in both bacteria and plants, causing an accumulation of lethal levels of ammonia in the cell (Tachibana *et al.*, 1986). Bialaphos resistant transgenic plants have previously been engineered for a number of species, including tobacco (De Block *et al.*, 1987), oilseed rape (De Block *et al.*, 1989), maize (Spencer *et al.*, 1990) and rice (Rathore *et al.*, 1993). These were all nuclear transformants, expressing a chimeric *bar* gene and were obtained by direct selection on regeneration medium containing PPT. Unfortunately, such a direct selection step was not possible for the generation of herbicide resistant tobacco plants expressing *bar* from the plastid genome. Instead, transplastomic plants were obtained by co-expression of the *aadA* marker gene and selection on spectinomycin. Once regenerated, the transplastomic plants were resistant to a maximum of 100 mg L\(^{-1}\) PPT and PAT accumulated to levels as high as 7%. The lack of recovery of any transplastomic clones by PPT selection is believed to have been due to inefficient inactivation of the herbicide by the initially transformed plastids (Lutz *et al.*, 2001).

### 1.5.3. Disease Resistance.

High-level expression of the antimicrobial peptide MSI-99 has been achieved by transformation of the tobacco chloroplast genome (De Gray *et al.*, 2001).
MSI-99 is an analog of magainin 2, which is a defense peptide, secreted from the skin of the African clawed frog (*Xenopus laevis*; Zasloff, 1987). Transplastomic plants were shown to be considerably more resistant to infection from phytopathogenic organisms such as *Pseudomonas syringae pv tabaci*, *Aspergillus flavus*, *Fusarium moniliforme* and *Verticillium dahliae*. Although its mode of action involves disruption of the microbial membranes (Huang, 2000), MSI-99 does not interfere with the chloroplast envelope or the thylakoid membranes. Transplastomic plants grew normally and produced seeds just like the wild-type control.

**1.5.4. Recombinant Protein Production in Plastids.**

Chloroplast transformation technologies have recently been extended to the production of medically useful proteins. The use of transgenic plants as production systems for therapeutic proteins has many potential advantages over current production methods e.g. pharmaceuticals recovered from plants would be free of contaminating microorganisms capable of causing disease in man. Human somatotropin (hST), produced in the pituitary gland, is currently being used in the treatment of dwarfism in children and may have benefits for those suffering from Turner syndrome and AIDS (Tritos and Mantzoros, 1998). Inserting the hST gene, complete with organelle-specific 5' and 3' expression signals into the inverted repeat region of the tobacco plastome has yielded some very encouraging results (Staub et al., 2000). The protein accumulated to levels as high as 7% total soluble protein, more than 300 times greater than the level achieved from a nuclear transgene described in the same publication. The plastid-expressed protein was also shown to be biologically active with the correct disulphide bond formation occurring. The latter step, which is inefficient in *E. coli*, may be facilitated in chloroplasts by the thioredoxin system (Ruelland and Miginiac-Maslow, 1999) or by a plastid localised protein disulphide isomerase (Kim and Mayfield, 1997).

The elastin-like protein polymer (GVGVP) has medical benefits in wound healing and tissue reconstruction in addition to non-medical uses as a biodegradable plastic (Urry et al., 1993; 1996). Attempts to express it in the chloroplast compartment via transplastomics gave disappointing results (Guda et
Although gene transcription occurred efficiently, driven by the strong \textit{rrn} promoter, western blot analysis revealed that polymer accumulation levels were up to 100-fold lower than anticipated. This discrepancy was postulated to be due to the plastid's glycine reserves being exhausted, a limitation that should be overcome by designing polymer genes that utilise amino acids present in higher concentrations in the chloroplast.

In addition, a recent scientific report (Mayfield \textit{et al.}, 2000) describes investigations, which are currently ongoing to engineer strains of \textit{Chlamydomonas} to express antibody genes in the chloroplast. Single chain antibodies that recognise epitopes of the herpes simplex virus, tetanus toxin and catalytic aldolase have been expressed from the plastome. So far however the antibodies have accumulated to low levels in the organelle and efforts are being made to increase them.

\textbf{1.5.5. Production of Edible Vaccines.}

The production of edible vaccines in plants through genetic modification is now considered a potential measure towards the control of many human and animal diseases (Giddings \textit{et al.}, 2000). Low-level expression of a number of transgenic antigens has been achieved from the nucleus (Mason \textit{et al.}, 1992; Haq \textit{et al.}, 1995; Takaberry \textit{et al.}, 1999), but may be insufficient to trigger an immune response when administered orally. Expression levels of the Cholera toxin B subunit (CTB) were greatly increased (4.1\% of total soluble protein) in transplastomic tobacco lines (Daniell \textit{et al.}, 2001). The chloroplast-synthesised CTB was antigenically identical to the native protein and could assemble into functional oligomers, suggesting the correct quaternary structure for the transpeptide.

Transplastomic edible vaccine technology has moved one step closer with the recent transformation of the tomato chloroplast genome (Ruf \textit{et al.}, 2001). Not only was high-level transgene expression observed in chloroplasts, it was shown to be almost 50\% higher in the chromoplasts of the fruit.

For certain biotechnological applications, chloroplast transformation possesses a number of characteristics that give it a superior edge over current nuclear transformation techniques.

1.6.1. Transgene Containment.

The tabloid press, particularly in Europe, and anti-GM crop activists like Greenpeace (see http://www.greenpeace.org) have presented transgenic-technologies as inherently unsafe to both the consumer and the environment. Although most of the allegations made are unfounded scare stories, safer, more "environment-friendly" alternatives are constantly being investigated. One of the major concerns surrounding the use of transgenic plants is the fear of transgene escape to wild relatives by pollen or seed dispersal. For example, the lateral transfer of genes, such as those conferring herbicide resistance could result in the generation of "superweeds", insensitive to high doses of various herbicides. Such an occurrence was shown to be possible at least in theory for sunflowers, strawberries and transgenic oilseed rape (King, 1996; Mikkelsen et al., 1996).

Transformation of the plastid genome is presently under consideration as a practical solution to controlling this dilemma. Plastids are strictly maternally inherited in most angiosperms (Corriveau and Coleman, 1988). Exceptions include alfalfa (Avni and Edelman, 1991) and occasionally tobacco (Medgyesy et al., 1986), both of which demonstrate biparental inheritance, but the phenomenon holds true for most commercially important crops (Daniell and Varma, 1998). The containment of a herbicide resistance transgene by transforming it into the plastid genome has been demonstrated (Daniell et al., 1998). However, the most concrete results obtained that support the transgene-confining reputation of transplastomics comes from a field study that looks at movement of a plastid marker gene from *B. napus* to its wild relative *B. napa* (Scott and Wilkinson, 1999). The study revealed no pollen-mediated plastid gene movement over a period of 3 years. Although the two species were capable of a sexual cross, the noted rapid decline and extinction of the hybrid progeny lines, minimises the risk of a possible introgression of the transgene into the weed species. For this to happen, the transplastomic plant would
have to act as the female parent and receive weed pollen over a number of generations. This report confirms that engineering the plastid genome of oil seed rape gives strong protection against transgene spread, but each crop species must be considered on an individual basis, taking into consideration factors like its mode of plastid inheritance and its ability to hybridise with related species. A widespread investigation has shown that of 60 cultivated crops examined, 49 have at least one closely related wild relative with which it can cross (Keeler et al., 1996).

Engineering a Bt toxin gene into the plastid genome of species where predominantly maternally inherited plastids are known (McBride et al., 1995; De Cosa et al., 2001) reduces the risk of insecticidal protein expression in transgenic pollen. This reduces a major portion of the risk of toxicity to non-target insects, such as monarch butterfly larvae, which under laboratory conditions have been shown to be harmed by ingestion of pollen expressing the toxin from the nuclear genome (Losey et al., 1999).

It has recently been proposed that differences in codon usage could also be utilised in the prevention horizontal transfer of transgenes from the plastid genome to bacteria (Lutz et al., 2001). The bacterial bar gene was altered (s-bar) so that its codons were more suited to high-level chloroplast expression, resulting in a reduction in the GC content from 67.9% to 45.8%. Compared with the wild-type gene, S-bar expression was greatly reduced in E. coli. Thus in the unlikely event of such a gene being transferred from the transplastomic plant to a bacterial host (Sylvanen, 1999), it would not be expressed sufficiently to provide a selective advantage on the organism and would ultimately be lost.

1.6.2. Polycistrionic “Operon” Expression in Plastids

Although the introduction of a single gene can alter some plant characteristics, the traits that are of most interest to breeders are controlled polygenetically. Eukaryotic nuclear genes are transcribed as single units and are not arranged in operons. It took a total of seven years work to introduce a simple three-gene biosynthetic pathway into rice by a succession of independent Agrobacterium transformation events (Ye et al., 2000). B-carotene was produced,
only after these transgenic plants were back-crossed so all three genes resided in one strain. Efforts were compounded by each cassette's transformation having to be screened for the best expressing strain because of the random nature of the *Agrobacterium* targeting.

Expression of the *Bt cry2Aa2* three-gene operon from the chloroplast genome of tobacco was much more straight-forward (De Cosa *et al.*, 2001). Only one transformation event was required as all three genes were transcribed from a single 16S rRNA promoter. Because homologous recombination ensures controlled targeting of the foreign gene(s) to an exact location in the plastid genome, position effect considerations are eliminated and large-scale screening programs avoided. The chloroplast also appears to be devoid of a lot of the epigenetic effects observed in the nucleus. Phenomena such as transcriptional gene silencing or induced mRNA decay often interfere with nuclear transformation experiments, leading to reduced or complete loss of transgene expression (Kooter *et al.*, 1999).

The introduction of another bacterial-type operon into the plastid genome of tobacco led to the synthesis of polyhydroxybutyrate (PHB) from acetyl-CoA (Nakashita *et al.*, 2001). PHB is a biodegradable polyester considered to be a possible alternative to petroleum-based plastics and has already been synthesized in leaf chloroplasts of *A. thaliana* by nuclear transformation, utilising transit peptide sequences to direct the biosynthetic enzymes to the organelle (Nawrath *et al.*, 1994; Bohmert *et al.*, 2000).

**1.6.3. High-level Expression.**

The elevated copy number of the chloroplast genome relative to the nuclear genome (>1,000:1) is probably the most important factor accounting for the fact that plastid transgene expression levels often exceed those achievable by nuclear transformation. The expression level achieved for the Cry2 insecticidal protein (46.1% of total soluble protein) in plastids, is the highest level of foreign gene expression reported in transgenic plants to date (De Cosa *et al.*, 2001). Moreover, there is currently a wealth of information about plastid gene expression signals (see
Section 1.2) so choosing appropriate promoter, terminator and ribosome binding-site sequences can be done in a logical manner.

1.6.4. Compartmentalisation of Expression.

The compartmentalisation of the transgene product within the plastid could also prove beneficial in a number of ways:

1. Protein-containment in the organelle may simplify purification of pharmaceuticals such as somatotropin (Staub et al., 1999), during large-scale production.

2. Certain foreign proteins may be more stable in the plastid environment than in the cytoplasm because of the more limited spectrum of protein degradation pathways found in plastids (Adam, 2000).

3. A protein may be less toxic to a plant cell if it is localised in an organelle where it is inert. It was recently reported that high-level expression of GFP in the cytoplasm could affect plant morphology and inhibit regeneration (Haseloff and Siemering, 1998). However, there appears to be no deleterious effects when it is expressed at even greater levels in the chloroplast (Siderov et al., 1999; Khan and Maliga, 1999).

1.6.5. Limitations and Future Directions.

The limited number of species that have been successfully transformed presently restricts the utility of transplastomic biotechnology. Thus far, chloroplast transformations are routinely achievable in only one higher plant species: tobacco. Recently however, the chloroplast genome of tomato (Lycopersicon esculentum L.) has been transformed and fertile transplastomic plants regenerated (Ruf et al., 2001). Both A. thaliana (Sikdar et al., 1998) and potato (Siderov et al., 1999) have been used to generate stable plastid genome transformants, but unfortunately all the transplastomic plants recovered were infertile (for reasons unknown). The chloroplast genome of rice has also been transformed but an inefficient selection and segregation system meant that only heteroplasmic plants were regenerated (Khan and Maliga, 1999). In order to increase the species repertoire and to include more economically important crops, advances will have to be made, by developing
existing and new techniques for DNA delivery, and for regenerating and selecting transformants.

1.7. Thesis Aims

The major aim of the research work described in this thesis was to facilitate the application of chloroplast transformation technologies in *Brassica* crop species, in particular in *Brassica napus* L. (oilseed rape). This required (a) a detailed molecular characterization of the chloroplast genome of *B. napus* (Chapter 2) and (b) the construction of *Brassica*-specific plastid transformation vectors (Chapter 3). Finally, in order to address continuing concerns about the use of bacterial antibiotic resistance genes as selectable markers in plastid transformation, the possibility of using a Cre recombinase/loxP-based excision system for marker gene removal in chloroplasts was also investigated (Chapter 4).
CHAPTER 2
THE PLASTID GENOME OF
BRASSICA NAPUS

2.1. Introduction.

2.1.1. Brassica napus (oilseed rape).

Oilseed rape is a crop of increasing commercial value. Its oil is used in industry and higher nutritional grades are used for human consumption. Historically it has been cultivated in India for more than 3,000 years and in Europe since the 13th century (Labana and Gupta, 1993). The 1950s saw the start of large-scale rapeseed production in Europe. Today, total world seed production exceeds 22.3 million metric tonnes per year. The B. napus plant produces yellow flowers (see Figure 2.1) that following fertilisation produce pods. Within the pods are tiny round seeds that are crushed to obtain the oil. Each seed is approximately 40% oil by weight. The remainder of the seed is processed into meal and used as high protein animal feed. Canola is a genetic variant of rapeseed developed by breeding techniques in Canada. It is considered to be nutritionally superior, due to its lower level of saturated fatty acids (7%).

Although the plant’s ability to form a dense canopy enables it to compete with most weeds, the use of the herbicides fluazitop-butyl (Fusilade 2000®) and glyphosate (Roundup®), in addition to tillage, provides supplementary weed control. Several infectious diseases can reduce the yield potential of the species, e.g., brown girdling root rot (see Figure 2.2), whitemould (Sclerotina sclerotiorum) and black spot (Alternaria raphani). Proper management practices can reduce the threat of disease, including one-in-four year crop rotation, sanitized fields, and disease-free certified seed. Several insect pests are also known to infest oilseed rape including flea beetle, cabbage seedpod weevil, armyworms and aphids. There are a number of registered pesticides available to control the insect problem, e.g., Dipel® for armyworm treatment (see Figure 2.2). Plant breeders are constantly
looking for new and better ways of providing growers with improved yield and quality.

Genetic modification technologies will soon be used to improve the nutritional content of many crop plants. It is well understood that all the enzymes required for fatty acid synthesis in plants are compartmentalised inside the plastid (reviewed in Slabas and Fawcett, 1992). Plastid genome encoded acetyl-CoA carboxylase (ACC) converts acetyl-CoA (entering the plastid as acetate) into malonyl-CoA. Fatty acid synthetase (FAS) uses these two substrates to make fatty acids up to 18 carbon residues in length and desaturation can occur by the action of soluble stearoyl-ACP desaturase. Fatty acids are then either incorporated into complex plastid lipids or undergo hydrolysis and are exported from the plastid. Further modifications may occur in the cytoplasm before they are esterified to a glycerol backbone. An enormous variety of triglycerides (storage lipids) can be generated by different fatty acid structural permutations (Hillditch and Williams, 1964). Understanding these biosynthetic pathways will in time lead to the implementation of both nuclear and plastid genome transformation technologies aiming to alter the lipid content and constituents of oil producing crops.

2.1.2. Evolution of the Angiosperms.

The angiosperms or flowering plants are unquestionably the most dominant plant group on earth. Therefore, it is surprising that until recently many basic questions surrounding their origins and rapid morphological diversification remained unanswered (Crane et al., 1995). Even Charles Darwin referred to the emergence and success of the angiosperms as a "perplexing phenomenon" and an "abominable mystery" (Darwin, 1903). Progress in reconstructing the overall phylogeny of angiosperms has been achieved using both fossil records (Friis et al., 2001; Zhou et al., 2001) and analysis of DNA sequences (Qiu et al., 1999; Soltis et al., 1999; Kuzoff and Gasser, 2000). The phylogenetic trees derived from these studies have a much more concrete basis than those traditionally derived from studies of floral morphology (Endress, 1994), reproductive biology (Weller et al., 1995) and life histories (Dodd et al., 1999).
The most widely accepted phylogenetic tree of the angiosperms is branched into a number of clades (Angiosperm Phylogeny Group, 1998). The most notable of these being the monocots (rice and maize), rosids (rose, Arabidopsis and B. napus) and asterids (tomato, spinach and tobacco). The asterids and rosids are members of a grouping known as the eudicots that diverged early in the radiation of dicotyledonous plants. Fossil evidence has shown that the asterids and rosids evolved from a common ancestor more than 90 million years ago (Gandolfo et al., 1998) and mitochondrial sequence analysis dates the divergence at 112-156 million years (Yang et al., 1999a). The divergence between the monocots and the eudicots was estimated to have occurred at 200 (± 40) million years, based on both plastid and nuclear DNA sequence analysis (Wolfe et al., 1989). This figure considerably predates the earliest appearance of monocots in the fossil records (Bremer, 2000).

The earliest fossils generally accepted as angiosperms are pollen deposits 132-141 million years old (Brenner, 1996). However, the time when angiosperms first diverged from sister groups such as the cycads (extant gymnosperms) is probably closer to 340 million years, an upward figure estimated using nucleotide substitution rates and a reliable internal calibration point (Wolfe et al., 1989). It has been postulated that early angiosperm radiation during the Late Jurassic to Early Cretaceous Period was influenced by the coevolution of insect pollinators (Crepet and Nixon, 1996). Indeed fossil evidence of the nectar-collecting Branchyera flies during the Late Jurassic Era implies flowering plants most likely originated during the Middle Jurassic (Ren, 1998). Elevated levels of atmospheric CO₂ and even changes in the behaviour of large herbivorous dinosaurs may also have played a critical role in the initial stages of angiosperm radiation (Barrett and Willis, 2001).

Recent phylogenetic analysis of nuclear 18S rRNA sequences from a wide range of seed plants revealed that angiosperms and gymnosperms form two independent monophyletic groups (Chaw et al., 1997). In other words their common ancestor is more likely to be an ancient fern than an extant gymnosperm. This contradicts earlier views based on cladistic analysis of morphological data suggesting that the ancestors of angiosperms were derived from the gymnosperm
lineage (Doyle and Donoghue, 1986). Basal angiosperm phylogeny is another point of recent debate. Qiu et al. (1999) describes the shrub *Amborella trichopoda* as being sister to all other angiosperms, possibly representing an early stage of angiosperm evolution. The next diverging lineage (next branch up the tree) includes the water lilies (Nymphaeales) followed by a branch consisting of the Austrobaileyacea, Trimeniaceae, Illiciaceae and Schisandraceae families. The conclusions of Barkman et al. (2000) disagree with this branching system, claiming instead that both *Amborella* and the Nymphaeales together make up the basal angiosperm branch. The significance of the later tree structure means that the earliest ancestors of angiosperms may have had either unisexual or bisexual flowers.

Initial molecular analysis of plant phylogeny compared the sequences of single genes, which could only yield weekly supported and often contradictory trees (Chase et al., 1993). Confidence in branching orders, as measured by bootstrap values (Felsenstein, 1985) or parsimony jackknifing (Farris et al., 1996), could be greatly improved using multiple genes as well as a larger taxa sampling number (Soltis et al., 1999). Including sequences from all three plant genomes (nuclear, mitochondrial and plastid) in the analysis is preferable as it would reduce genome-specific evolutionary biases such as rate heterogeneity, GC content bias and RNA editing (Qiu et al., 1999). Detailed studies like these have been made possible with the help of advancing computer and software technology (Kei and Nei, 2000).

2.1.2.1. Evolution of the *Brassica* Lineage.

Although the Brassicaceae family lineage includes a number of important crop species, it has not received the level of molecular analysis it perhaps deserves. The family comprises more than 3000 species in approximately 350 genera and is distributed worldwide (Price et al., 1994). The model plant *A. thaliana*, whose entire nuclear, chloroplast and mitochondrial genomes have recently been sequenced (The *Arabidopsis* Genome Initiative, 2000), also belongs to this family. This sequence data should provide a valuable resource for the identification of homologous *Brassica* genes suitable for biotechnological manipulation (Paterson *et
al., 2001). Comparative mapping of the nuclear genomes of *Arabidopsis* and various *Brassica* species supports the hypothesis that the increased size of the *Brassica* genomes is as a result of extensive genome duplication (Lagercrantz, 1998), but shows that the gene order is largely conserved (Cavell *et al*., 1998). The genus *Brassica* is polyploid, with diploid species such as *B. rapa* L. (genome AA) and *B. oleracea* L. (genome CC) capable of hybridising to form a tetraploid species like *B. napus* (genome AACC; Paterson *et al*., 2001).

Yang *et al.* (1999b) used the sequence of the internal transcribed spacer region (ITS) of the 18S-25S nuclear ribosomal DNA to elucidate the phylogenetic relationships between *Brassica* and related genera including *Arabidopsis*. They concluded that *Brassica* species were divided into two evolutionary lineages: the *nigra* lineage and the *rapa/oleracea* lineage. This phylogenetic study also concludes that the *Brassica* and *Arabidopsis* lineages diverged prior to the *Brassica-Sinapis* and *Brassica-Cardamine* splits. This leads to the question of when exactly did the *Brassica-Arabidopsis* split occur. Fossil evidence suggests a date of at least 10 million years ago (Muller, 1981). More recently Yang *et al.* (1999a) performed molecular phylogenetic analyses based on the nucleotide sequence of the first intron of the mitochondrial gene for NADH subunit 4 (*nad4*). The study included 18 species from Brassicaceae, two monocots (maize and wheat) and an asterid. A previously obtained estimate of the maize-wheat divergence at 50-70 million years (Wolfe *et al*., 1989) was used as a fixed point reference to calculate a nucleotide substitution rate of $0.16-0.23 \times 10^{-9}$ substitutions per site per year for the 1.4kb mitochondrial sequence. The parameters of the tree were then used to estimate the dates of divergence between monocots and dicots to be 170-235 Myr; between rosids and asterids to be 112-156 Myr and between *Brassica* and *Arabidopsis* to be 14.5-20.4 Myr. However, the authors do acknowledge that the estimates must be treated with caution on account of the limited sequence data used in their analysis (Yang *et al*., 1999a).

### 2.1.3. Construction of *B. napus* Plastid Genome Libraries.

The elucidation of the genetic map and subsequently the entire nucleotide sequence of the plastid genome of tobacco (*Nicotiana tabacum*) was facilitated by
the construction of a clone bank or library (Sugiura et al., 1986). This comprised a cloned set of chloroplast DNA restriction endonuclease fragments, which covered the entire genome. This basic strategy has since been used to obtain the entire nucleotide sequence of the chloroplast genomes of liverwort (Ohyama et al., 1986), rice (Hiratsuka et al., 1989), pine (Wakasugi et al., 1994) and maize (Maier et al., 1995). In the case of the Arabidopsis plastid genome (Sato et al., 1999), DNA templates were isolated from a phage P1 library prepared from whole cellular DNA (Liu et al., 1995). In contrast, the Oenothera elata plastid genome was sequenced from PCR-derived DNA fragments as well as restriction fragment clones (Gordon et al., 1982; Hupfer et al., 2000). Restriction fragment clone banks also serve as very useful primary DNA sources for the construction of plastid transformation vectors (Svab et al., 1990).

This chapter describes (a) the construction and detailed characterization of B. napus plastid genome libraries, (b) sequence analysis of approximately one-third of the plastid genome and (c) phylogenetic analyses using sequence data from B. napus and seven other species.
2.2. Materials and Methods.

2.2.1. Materials.

2.2.1.1. Plant Materials.

Seeds of the *Brassica napus* cultivar Licosmos, were supplied by Teagasc (http://www.teagasc.ie) and grown under glasshouse conditions.

2.2.1.2. Bacterial Strains and Plasmids.

*Escherichia coli* DH5a [reoR, endA1, gyrA96, hsdR17 (rKmK'), recA1, relA1, supE44, thi-1, Δ(lacZYA-argFV169), f80lacZ ΔM15, F−, Γ] (Hanahan, 1983) was used as the host for cloning experiments and for the amplification of plasmid DNA for purification.

pIC19H (Marsh *et al.*, 1984) was used for the cloning of the *B. napus* chloroplast genome *HindIII* restriction fragment library.

pUC7 (Messing and Vieira, 1982) was used for the cloning of the *B. napus* chloroplast genome *SalI* and *EcoR*I restriction fragment libraries.

pUC19 (Yanisch-Perron *et al.*, 1985) was used for the cloning of the *B. napus* chloroplast genome *XbaI* restriction fragment library.

2.2.1.3. Enzymes.

Restriction endonuclease enzymes were purchased from New England Biolabs, Boehringer Mannheim, Promega and GibcoBRL. T4 DNA ligase, the Klenow fragment of DNA polymerase I and alkaline phosphatase were purchased from Boehringer Mannheim. T4 polynucleotide kinase (PNK) was purchased from New England Biolabs. *Pfu* DNA polymerase was purchased from Stratagene. Radioactive nucleotide triphosphates and Hybond–N blotting membranes were purchased from Amersham Life Science.

2.2.1.4 *E. coli* Growth and Selective Media.

*LB* (Luria-Bertani) broth contained 10g/l Bacto-tryptone, 5g/l Bacto-yeast extract, 5g/l NaCl, pH 7.0 and supplemented with 1.5% agar for solid medium;
ampicillin (100mg/L) or kanamycin (50mg/L) were added to select for plasmid containing cells; isopropyl β-D-thiogalactopyranoside (IPTG; 0.025mM) and 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal; 0.005% w/v) were added to media to facilitate identification of insert containing pIC19H transformants.

2.2.1.5. Oligonucleotides used for DNA Sequencing.

All PCR cycle-sequencing reactions were carried out using the fluorescent dye terminator kit supplied by Amersham based on the Sanger dideoxy method (Sanger et al., 1977). Reactions were performed in accordance with the manufacturer’s instructions and analysed on the Applied Biosystems model 373A automated sequencer. All synthetic oligonucleotides were manufactured by MWG-Biotech AG and received in lyophilised pellet form.

To position the clones of the Brassica chloroplast genome library relative to existing maps the ends of the cloned chloroplast DNA fragments were sequenced using the M13/pUC primers:

M13 forward: 5'-GTT TTC CCA GTC ACG AC-3'
M13 reverse: 5'-AGC GGA TAA CAA TTT CAC AC A GGA-3'.

To complete the 9419 bp nucleotide sequence of the 16S rRNA gene region, two HindIII clones (Hin 34 + Hin 18) were used as the template and a "primer walking" strategy was employed by designing primers from the end of existing sequences. In addition to those sequences obtained using the M13/pUC primers, additional “starting points” were created by designing primers based on closely related chloroplast DNA sequences of N. tabacum (Shinozaki et al., 1986) and A. thaliana (Sato et al., 1999) in regions expected to be conserved (coding sequences). Primers designed on the basis of such similarity are underlined in the following list:

98863f: 5'-CAT TCC TCC TAG AGT AGC TG-3'
99078f: 5'-CGA TCA TCA GAA GAA GGA TTA GGC C-3'
99426f: 5'-GAG AGG GCT TAG TTG ATC CAT G-3'
99647f: 5'-GGC AGC ATC CAC TAA TTC GGA AC-3'
99968f: 5'-CGA TTA CGA TAA ATT GGA TCG G-3'
100464f: 5'-CTT TTG ACA TAT AAG AGT TCC G-3'
To complete the 11002 bp sequence in the *rbcL* region, incorporating two *HindIII* clones (Hin 32 + Hin 27), the following primers were used:

- **ndhfl**: 5'-GCA GAA ACA TAG ACG AAC TCC-3'
- **ndhO**: 5'-ATT TCT GTT TCT TAA ATT TGC TTT ATG-3'
- **ndhfS**: 5'-GAA CCA ACA AAT TGA AAG TAA CCA G-3'
- **ndhf6**: 5'-GAA ATG GTT CCA AGT ACT CTG-3'
- **tm vfl**: 5'-CCT AGC ATT GAG TAG ACT GGG TA-3'
- **tmvfZ**: 5'-AGC CCT TGT GTT TAT GAT CCA C-3'

These primers were used in conjunction with two *HindIII* clones (Hin 32 + Hin 27) to complete the sequence.
To complete the 36132 bp sequence in the Large Single Copy region, spanning from \textit{rps4} to \textit{atpA}, the following primers were used:

\textbf{BA1:} 5'-GGT TCA AAT CCT ATT GGA CGC-3'
\textbf{BA1b:} 5'-CAT AGG TCT TCT AAA CCT TTG G-3'
<p>| BA2: 5'-CAA AAA GCT TCT GCT TCA GCG G-3' |
| BA3: 5'-GAA AAT TGG GAA AAG GCT TCT AAT TC-3' |
| BA4: 5'-AGA GCT TAA TTT AGC GGC TCT TTC-3' |
| BA5: 5'-CCT GAG CCA CGG AAG AAG CTT TT-3' |
| BB1: 5'-CCA TCA ATA GGG TTA GCC AAG GC-3' |
| BB2: 5'-GCT CAA TAC GTT CAC GGA TAA TA-3' |
| BB3: 5'-GTA AAG CTT GGT GAA AAA CCC G-3' |
| BC1: 5'-CAC GCA GTT CTT CCG AAT TTC G-3' |
| BC2: 5'-CTA TTT GGA GGA TTC TTC TGA CC-3' |
| BC3: 5'-GAG AAT CAA AGT CAA GTA GAC TTG TTA CC-3' |
| BC4: 5'-GTT CCA ATA GGA GAA CCC AGC CC-3' |
| BC5: 5'-GCT AGA ATT TGA GAC TTG TTA ATG TC-3' |
| BC6: 5'-GCT TCC ATA AAA GCC AAA CTA AGC-3' |
| BE1*: 5'-GTT CCA ATA GGA GAA CCC AGC CC-3' |
| BE2: 5'-GAT CGT TTT TTT CAG CCA ATT CAC AG-3' |
| BE3: 5'-ATC CAA GGA ACA TGA CAG GTA TAG-3' |
| BE4: 5'-TTG CTT GGT GCT GCT AAG TCC C-3' |
| BF1: 5'-GCC AAT AGA AAT GGT GGC GTA C-3' |
| BF2: 5'-ATA TAG CTA GAA CGA CCC TCA-3' |
| BF3: 5'-CAA GTC CCT AAA CTT ATG AAG TCT T-3' |
| BF4: 5'-CCA AAA TGA ACT CCT GCT CTG ATC-3' |
| BF5: 5'-CTC TCA TAT CTC CCT CTA ATA AAG AT-3' |
| BF6: 5'-GCA ACT AAT TGA ATA CTT GTG TCT GC-3' |
| BG*: 5'-CGA ACT AAT TGA ATA CTT GTG TCA G-3' |
| BG*:2: 5'-CGC CGA CTT GGC TTC ATG ATC C-3' |
| BG*:3: 5'-CCG TTT TCA TCT ATT AAA TAA GAG TTG-3' |
| BG*:4: 5'-CTA AGG GAT TGT TGA AAG ATG CC-3' |
| BH*: 5'-GAC ATT AAT CCT CTC ATA CCT ACT A-3' |
| BH*:2: 5'-ATG AAA TGG ATG TAG CAG TTG GCT G-3' |
| BH*:3: 5'-ATT CAC ACG CGT TGA TCT AAT TGC C-3' |
| BH*:4: 5'-GCA ATA ACG CGT TGA TCT AAT TGC C-3' |
| BH*:1: 5'-CTA CCG AAA ATA CCA TTC ATG GG-3' |
| BH*:2: 5'-CTG AAT AAT TGG GAA TTC CAG ATT-3' |
| BH*:3: 5'-AAA GTG GAT GTT CCC TCG CGA AT-3' |
| BH*:4: 5'-CAA TAG ATT GAC CGC CAA TAA TAC C-3' |
| BH*:5: 5'-GGA TAA AAA TTC TTT CTG ACA TCA TCC-3' |
| BI*: 5'-GCC GAT AAG CTC CAA TTG CAT C-3' |
| BI*:1: 5'-GCC ATT TGA CTC CCA TCA AAG-3' |</p>
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI3</td>
<td>5'-TCT CGC GAA ATC TTC CCT CTT-3'</td>
</tr>
<tr>
<td>BI4</td>
<td>5'-AAA TGC TTA GCT AAT TCC ATA CGT C-3'</td>
</tr>
<tr>
<td>BI5</td>
<td>5'-ATA GGG TCC ACT ATA TGA TCT CC-3'</td>
</tr>
<tr>
<td>BI6</td>
<td>5'-AAT TCA TAC GGG ATT TAT TTG TCT ATG-3'</td>
</tr>
<tr>
<td>BJ1</td>
<td>5'-CTG AAT CGA CCA TAG ACC AAT CG-3'</td>
</tr>
<tr>
<td>BJ2</td>
<td>5'-CCA CAT TGT TCA CAA AAT TGC CAG-3'</td>
</tr>
<tr>
<td>BJ3</td>
<td>5'-CAG AGC CAA AGA TCG TAG TTC T-3'</td>
</tr>
<tr>
<td>BJ4</td>
<td>5'-CAG ATA CAA GGA AAT GAT TCA GTT CC-3'</td>
</tr>
<tr>
<td>BJ5</td>
<td>5'-CCT GAT GAA TTA ATT TCA AGA TAT AGG G-3'</td>
</tr>
<tr>
<td>BJ6</td>
<td>5'-TCC TGT CTA GGC AAA ATT TTG GAA-3'</td>
</tr>
<tr>
<td>BK1</td>
<td>5'-GAT CCT AGC ATC ACA ATT CC-3'</td>
</tr>
<tr>
<td>BK2</td>
<td>5'-GTT TGA GCG TTG ATA CAT AAT TAA TGG-3'</td>
</tr>
<tr>
<td>BK3</td>
<td>5'-CCT CTT CCC ATG CGA TAG TCA-3'</td>
</tr>
<tr>
<td>BK4</td>
<td>5'-CCA TGA ACG ATT TGT GTC AAT GGA-3'</td>
</tr>
<tr>
<td>BK5</td>
<td>5'-AGT CTC CAG TTA ATA TTT CGG CG-3'</td>
</tr>
<tr>
<td>BL1</td>
<td>5'-CGA ACC CAT AGC TGA TAG-3'</td>
</tr>
<tr>
<td>BL2</td>
<td>5'-CAG GTT CTA CCA ATT GAT ATG TTT CC-3'</td>
</tr>
<tr>
<td>BL3</td>
<td>5'-CAC ATG AAA TTT TAT CCA ACT CCA AG-3'</td>
</tr>
<tr>
<td>BL4</td>
<td>5'-GGA TCT GTT CTA GCA GAT AAA C-3'</td>
</tr>
<tr>
<td>BL5</td>
<td>5'-TGC ACA GGT ATA TAT ATA TGT CTC-3'</td>
</tr>
<tr>
<td>BL6</td>
<td>5'-GTG TTA TAA TTT AAA TGG AAG AAG ATT TC-3'</td>
</tr>
<tr>
<td>BM1</td>
<td>5'-CAT GGC CGA GTG GTA AGG C-3'</td>
</tr>
<tr>
<td>BM2</td>
<td>5'-CGA TAT AGA CTC ATG AAA GTG TCG-3'</td>
</tr>
<tr>
<td>BM3</td>
<td>5'-CTC TGC ACC AGT GAT TTC ACT AT-3'</td>
</tr>
<tr>
<td>BM4</td>
<td>5'-GTT AAC CAT TCA TGG GAT AAC CC-3'</td>
</tr>
<tr>
<td>BN1</td>
<td>5'-GCC CCT TAG GAT ACT ACT GAA TC-3'</td>
</tr>
<tr>
<td>BN2</td>
<td>5'-CAA TAT CGG AGC TAT AAA ATC GAT TC-3'</td>
</tr>
<tr>
<td>BN3</td>
<td>5'-GAG AAA ACA AGT ACG AAA TTT CAA TTT G-3'</td>
</tr>
<tr>
<td>BN4</td>
<td>5'-AAA GAG GAA ACC CCA TAC AGT ATC-3'</td>
</tr>
<tr>
<td>BO1</td>
<td>5'-AAC CCG CAG CTT CCG CCT T-3'</td>
</tr>
<tr>
<td>BO2</td>
<td>5'-GGA AAC AAA AAG TAA CGA GCC AC-3'</td>
</tr>
<tr>
<td>BO3</td>
<td>5'-CAT ACT TGC TCA ACC GCC ATC AT-3'</td>
</tr>
<tr>
<td>BO4</td>
<td>5'-ACT TAA TAG ATT GCC TCA TCT CAT TG-3'</td>
</tr>
<tr>
<td>BP1b</td>
<td>5'-GGC CCT TTT AAC TCA GTG GTA G-3'</td>
</tr>
<tr>
<td>BP2</td>
<td>5'-TGA ATC GCC AAA TAG ATA TCC GTT G-3'</td>
</tr>
<tr>
<td>BP3</td>
<td>5'-ATA TAT TGC GTT CCT GCC TAG AGA-3'</td>
</tr>
<tr>
<td>BP4</td>
<td>5'-TAA CCG GCA TGG ACA TGT TGA TAT-3'</td>
</tr>
<tr>
<td>BP5</td>
<td>5'-CGC TAT GAC TAT AGC CCT TGG T-3'</td>
</tr>
<tr>
<td>BP6</td>
<td>5'-CCG CTG CAG TTT CTA CTC CTG C-3'</td>
</tr>
<tr>
<td>BQ2</td>
<td>5'-CGT CAA TTT GAA CTT GCT CGA TC-3'</td>
</tr>
</tbody>
</table>
BQ3: 5'-TTA ACC CAA CTC AAG CCG AAG-3'
BQ4: 5'-TTC TAC CAC GTG GAA ACG CTC TTT AA-3'
BQ5: 5'-TTG GGC TTT GGC GGT ATT TAT C-3'
BR*: 5'-TCT GG TAC TCT TTA TAT GGA T-3'
BR*2: 5'-CTT CTG TAA CTT GAG TGA AAT ATG AAT C-3'
BR1: 5'-GGA TGG ATT GTT AGT GTG GAC G-3'
BR2: 5'-AGA CCA ACG TCT TGG AGC TAA C-3'
BR3: 5'-GTC TCT CCG AGA AGT TGG TTA TC-3'
BR4: 5'-GTC TCT TTT TTT TTA ATT CAA TCT CT-3'
BR5: 5'-GGG TGT GAC TTT TTT TAT TAT AGA TCT ATC-3'
BR6: 5'-TAT TTT CGT CTC CTG ATG GGT GGT-3'
BS1: 5'-TGC GGA TAT AGT CGA ATG GTA A A-3'
BS2: 5'-GGC ATG AAA CAA AAA AAC TTG GAC T-3'
BS3: 5'-TTT TCA CTT AGC GAC GGA ATC T-3'
BS4: 5'-GTC TAT GAA GCC AGA TGG TTA CAA G-3'
BT1: 5'-CCG TTT AGA TCT TAA TTT ATG ATC G-3'
BT2: 5'-CGT GCA TCT AAA GCC AAT AGC ATC C-3'
BT3: 5'-AAG TCC GAA ATG AGG GAA TTA C-3'
BT4: 5'-GGG TTT TGA GCA TAC AGA TTC C-3'
BU1: 5'-GCG ATG AAA CAA AAA AAC TTG GAC T-3'
BU2: 5'-CCG TTT AGA TCT TAA TTT ATG ATC G-3'
BU3: 5'-GAG CCC AAA CAA TGG ATT CAA TAA G-3'
BU4: 5'-CCA ACT GAA ATG GAA TAT TAC TAC CG-3'
BU5: 5'-CTT GTT GTA ATT GTA TAG CAG TAT C-3'
BV1: 5'-GCA CCA ACT ATG AGA AAT CCA CC-3'
BV2: 5'-GTT GGT CCT ATA CAT ATG ACC TG-3'
BV3: 5'-AAG GGA CCC AAG TCC TAG TA-3'
BV4: 5'-GAA AGA TGA TAG AGA GTT GGC G-3'
BW1: 5'-ACT TCT GGT TCC GCC GAA CG-3'
BW2b: 5'-CCA TAG GGT GCT CAA CGG ACC-3'
BW3: 5'-GTA CTA CTT ATC CCT AGA GAA TAC C-3'
BW4: 5'-AAT CTC CTT GTT GAA TGG CCT GTT-3'
BW6: 5'-GCA TGT CCA CTG TTA TTA GAT ATT AG-3'
BW7: 5'-TCT TTA CGG TGC TTC TCT TAT C-3'
BW8: 5'-CGT AAT GAC AGA TCA CAG CC-3'
BW9: 5'-GTC GCA TAG CTT CAT AAT TCT G-3'
BX1: 5'-GGG CTC CTC CTT TAG GTG CAT 3'
BX2c: 5'-ACG AGA TGG ATG TTT ATT GTT CCA A-3'
BX3: 5'-AAA CTT CAC CTT CAA AAG CGT AGA-3'
BX4: 5'-CGA ACT AAC ATA AAA GCG GAC TCG TT-3'
BY1b: 5'-TTG TTT ACC GAG GGT TCG AAT C-3'
BY2: 5'-CGA GAG TAA TAT TCT ACG ACA AGC-3'
BY2b: 5'-AAC TTG GTA TAT CAA CTA TAC GAC C-3'
BY5: 5'-GGA AAA AGG TGA AAG AAG TCT TTT C-3'
BY6: 5'-GGA AAA AGG TGA AAG AAG TCT TT-3'

2.2.2. Methods.

2.2.2.1. Molecular Cloning Techniques.
Nucleic acid manipulations were performed as described by Sambrook et al. (1989) unless otherwise stated. Restriction digests, alkaline phosphatase, T4 polynucleotide kinase and ligase reactions were carried out in accordance with the manufacturers’ recommendations.

2.2.2.2. Purification of Chloroplast DNA.

*Brassica napus* chloroplast DNA was purified using a modified method originally described by Milligan (1989). In this procedure crude intact chloroplasts were isolated, then lysed and their DNA purified. Plant material was kept at 4°C for the intact chloroplast isolation part of the procedure and at room temperature for the remaining steps unless otherwise stated. Approx. 60g of healthy leaf tissue was detached and homogenised in 300ml cold isolation buffer (1.25M NaCl, 50mM Tris, pH8.0, 5mM EDTA, 0.1% bovine serum albumin (w/v), 0.1% β-mercaptoethanol (w/v)) using a Moulinex blender. This was then filtered through four layers of cheesecloth. The filtrate was centrifuged for 10 minutes at 3000g in a Sorvall RC-5B centrifuge using a prechilled GSA rotor to pellet chloroplasts. The chloroplast pellet was resuspended in 30ml cold isolation buffer and repelleted by centrifugation at 3000g for 5 min. The final chloroplast pellet was resuspended in 10ml cold isolation buffer. To lyse the chloroplasts, 2ml (1/5 volume) of 10% CTAB (10% hexadecyltrimethylammonium bromide in H₂O (w/v)) was added. The mixture was incubated in a waterbath at 70°C for 20 minutes. An equal volume of chloroform was added and mixed gently but thoroughly. The mixture was clarified by centrifugation at 1500g for 10 min. The upper, aqueous phase was recovered and 2/3 volume of cold isopropanol was added to precipitate the DNA at -20°C overnight. The solution was then centrifuged at 1500g for 10 minutes to
pellet the nucleic acids and the pellet washed with 80% ethanol. The dried pellet was dissolved in 600μl TE buffer (10mM Tris, pH 8.0, 1mM EDTA). The cpDNA was reprecipitated by adding 1/10 volume 3M sodium acetate and 2/3 volume 100% ethanol. The solution was centrifuged at 12,000 rpm for 2 minutes and the final pellet resuspended in 600μl TE buffer.

2.2.2.3. Preparation of E. coli Competent Cells and Transformation.

DH₅α E. coli competent cells were prepared and transformed using a modified version of the procedure described in Mandel and Higra (1970). A single isolated colony was used to inoculate 10ml of LB broth and was grown overnight at 37°C. 1ml of this culture was then added to 50ml of LB broth and incubated at 37°C until the O.D.₆₀₀nm reached 0.5. The culture was then centrifuged for 10 minutes at 3000g in a Sorvall RC-5B centrifuge using a prechilled GSA rotor. The bacterial cell pellet was then resuspended in 20ml of 0.1M CaCl₂ and left on ice for 15 minutes. Cells were then pelleted by centrifugation for 5 minutes at 3000g and resuspended in 2.5ml of chilled 0.1M CaCl₂, 20% glycerol and stored at −70°C until required for transformation.

Typically 10-50ng of vector DNA from a ligation reaction was mixed with 70-100μl of competent cells and incubated on ice for 15 minutes. Heat shock was performed at 42°C for 1 minute and samples were returned to ice for 1 minute. Five volumes of LB broth (no antibiotic) was added and the mixture was incubated at 37°C for one hour followed by plating on LB agar containing the appropriate antibiotic.

2.2.2.4. Purification of Plasmid DNA (Boiling Method).

This plasmid “mini-prep” procedure is a variation of the method described by Holmes and Quigley (1981). An isolated colony of E. coli was used to inoculate 10ml of LB containing the appropriate antibiotic and incubated at 37°C and 200rpm, overnight. Cells were centrifuged at 3000g at 4°C for 10 minutes and the pellet resuspended in 60μl of TS buffer (50mM Tris-HCl, pH 8.0, 10mM EDTA, pH 8.0, 25% sucrose and transferred to Eppendorf tubes. 20μl of lysozyme solution (10mg/ml lysozyme in 250mM Tris-HCl, pH 8.0) was added and the cells
were allowed to lyse on ice for 15 minutes followed by the addition of 500μl of MSTET solution (5% sucrose, 50mM Tris-HCl, pH 8.0, 50mM EDTA, pH 8.0, 5% Triton X-100). The sample was boiled for 1 minute and then centrifuged in an IEC Micromax at 13,000rpm for 15 minutes. The “gluey” pellet containing sedimented chromosomal DNA and other cellular debris was removed using a sterile toothpick and discarded. The supernatant was treated with 7μl of RNaseA (10mg/ml) at 42°C for 20 minutes, followed by a single phenol extraction step. Plasmid DNA was precipitated by adding 2/3 volume of isopropanol, holding the mixture at -20°C for 10 minutes before centrifuging at 13,000rpm for 10 minutes. The pellet was washed once with 70% ethanol and the final pellet resuspended in 40μl of deionised water.

### 2.2.2.5. 32P Labeling of Chloroplast DNA Digests, Plasmids and Isolated DNA Fragments.

DNA probes incorporating α-[P32]-dCTP were generated by random priming as described by Feinberg and Vogelstein (1983). A 13μl aliquot of DNA solution (1.5ng/μl DNA fragment or plasmid; 25ng/μl digested chloroplast DNA) was denatured by boiling for 5 minutes and then placed on ice. A 13μl reaction mixture containing 400mM HEPES pH7; 40μM of each dGTP, dATP, dTTP; 10U/ml hexamer (random primer p(dN)6); 40mM β-mercaptoethanol; 100mM Tris-HCl pH8.0; 10mM MgCl2 was added along with 1μl of Klenow fragment of DNA polymerase I (2U/μl), 1μl BSA (10mg/μl) and 1μl α-[P32]-dCTP (370kBq/μl). The reaction was allowed to proceed at room temperature for 4 hours before being spun through a TE-equilibrated Sepharose CL6B mini column to remove salts and unincorporated label (Murphy and Kavanagh, 1988).

### 2.2.2.6. Transfer of DNA from Agarose Gels to Nitrocellulose Membrane.

DNA was transferred from agarose gels onto hybridisation membranes as described by Southern (1975). The DNA containing gel was gently shaken for 20 minutes in Denaturing solution (1.5M NaCl, 0.5M NaOH). It was then washed with Neutralising solution (1.5M NaCl, 1mM EDTA, 0.5M Tris-HCl pH7.2), shaking gently for 20 min. The transfer of the DNA to the membrane was allowed.
to proceed overnight by capillary action using 10X SSC solution (1.5M NaCl, 150mM Sodium Citrate, pH7.0) as the transfer buffer. Membranes were allowed to air-dry before the DNA was chemically fixed to the membrane by placing it on a UV transilluminator for 5 minutes.

2.2.2.7. Screening of Recombinant Bacteria by Colony Blotting.

Half of each bacterial colony was transferred to fresh LB-agar (including antibiotic) plates, positioned and numbered according to a grid. The other half of each colony was transferred to a duplicate plate with a Hybond-N membrane placed on the surface. Both plates were incubated overnight at 37°C. To release the DNA from the colonies, the membrane was placed, colony side up, on 3MM Whatman filter paper soaked in Denaturing solution (1.5M NaCl, 0.5M NaOH) for 15 minutes at room temperature and then transferred onto filter paper soaked in Neutralising solution (1.5M NaCl, 1mM EDTA, 0.5M Tris-HCl pH7.2) for 15 min. at room temperature twice. The membrane was washed by transferring it onto filter paper soaked in 2X SSC solution (0.3M NaCl, 30mM Sodium Citrate, pH7.0) for 15 min. The membrane was allowed to air-dry and the DNA was chemically fixed by placing it colony side down on a UV transilluminator for 5 minutes.

2.2.2.8. Hybridisation of DNA-Containing Membranes.

The procedure used was exactly the same for both colony blots and Southern gel blots. Prehybridisation was carried out at 65°C for 4 hours in the hybridisation buffer (0.9M NaCl, 90mM Sodium Citrate, pH7.0, 2g/L Ficol 400, 2g/L PVP 400K, 2g/L BSA, 0.1% SDS, 20µg/ml heat-denatured salmon sperm DNA). The radioactive probe was denatured at 100°C for 5 min., placed on ice for 1 min. and added to the hybridisation buffer. Hybridisation was carried out at 65°C overnight. The following day the membrane was washed initially in 2X SSC, 0.1% SDS for 20 min. at 65°C. The final stringent wash was carried out in 0.2X SSC, 0.1% SDS for 15 min. at 65°C. Having briefly dried the membrane it was then placed in a sealed plastic bag and autoradiographed.
2.2.2.9. Sequence Data Assembly and Analysis.

The sequence analysis program GeneJockey II from BIOSOFT® (www.biosoft.com) was used to assemble nucleotide sequence derived by primer walking. The genes and coding regions were identified by direct alignment with the genes found in the *A. thaliana* plastid genome (Sato *et al.*, 1999; Ac. AP000423 in the DDBJ/GenBank/EMBL DNA database). Any unidentified reading frames were subjected to a similarity search against the non-redundant protein database, using the BLASTP program (Altschul *et al.*, 1990).

A set of 31 plastid genome-encoded protein coding genes (exons only) and rRNA genes, from 7 completely sequenced plastid genomes were used in pair-wise alignments at both the nucleotide and the amino acid level, with the corresponding *B. napus* sequences. The protein coding genes were translated directly from the DNA sequence. Percentage similarities between pairs of sequences were estimated by calculating the number of positive identities between them, as identified by the default settings provided by BLAST from NCBI (www.ncbi.nlm.nih.gov/BLAST/). Twenty six of these protein sequences, which were common to all eight species, were concatenated end to end as described by Martin *et al.* (1998). A multiple sequence alignment between the concatenated sequences was performed using the CLUSTALW method (Thompson, *et al.*, 1994). This protein sequence alignment formed the basis for phylogenetic tree construction using the neighbor-joining method (Saitou & Nei, 1987; software provided by PhyloDendron, by D.G. Gilbert version 0.8d (iubio.bio.indiana.edu/treeapp/treeprint-form.html)), with distances estimated using Kimura's method (Kimura, 1980). Bootstrap values were calculated using PAUP* version 4.0 (Swofford, 1998). Sequence analysis and gene construct graphics were provided by the Vector NTI software package (www.informaxinc.com).
2.3. Results.

2.3.1. Construction of *B. napus* plastid genome libraries.

*HindIII* was chosen as the most suitable restriction enzyme for use in library construction. This decision was based on the size distribution of the restriction fragments produced by *HindIII* digestion of purified chloroplast DNA in comparison with those produced by other enzymes. Figure 2.4 shows the restriction band patterns obtained when samples of *B. napus* chloroplast DNA were digested with the enzymes *HindIII*, *EcoRI*, *SalI*, *BamHI*, *XbaI*, *XhoI* and *KpnI*. The majority of *HindIII*-digested fragments were considered sufficiently small (less than 10kb) to be easily cloned into a pUC based plasmid vector, yet sufficiently few in number (approximately 30) so that total clone coverage of the plastid genome could be achieved.

*HindIII*-digested chloroplast DNA was ligated into the cloning vector pIC19H. Following *E. coli* transformation, insert-containing colonies were identified on the basis of blue-white selection on agar containing X-Gal (DNA fragment insertion interrupts the β-galactosidase alpha-complementing gene and the resulting colonies are white on agar containing X-Gal; Marsh et al., 1984). Plasmid DNA minipreps were screened by digestion with the cloning enzyme *HindIII* in order to distinguish between different sized inserts. Then as a means of differentiating between similarly sized fragments, the clones were digested with both *HindIII* and *BamHI* on the basis that different chloroplast *HindIII* inserts of the same size would have different *BamHI* band patterns when electrophoresed on agarose gels. Figure 2.5 is of a picture of an agarose gel showing 23 unique *HindIII* clones (named by the prefix Hin followed by a number) digested with the cloning enzyme (*HindIII*). The various sizes of the chloroplast DNA fragments ranged from approximately 12kb to 1kb. The clone Hin126 contained two non-homologous chloroplast DNA fragments inserted into the vector’s cloning site, but all other *HindIII* clones contained a single chloroplast DNA fragment. The precise location of individual clones relative to the plastid genome of *A. thaliana* was then mapped by obtaining the end sequences of the inserts using the vector-based
primers M13 Forward and M13 Reverse. These end sequences were aligned firstly with the chloroplast genome map of N. tabacum (Ac. Z00044) and later with that of A. thaliana (Ac. AP000423). Table 2.1 describes the positional alignment of the end sequences for each of the cloned HindIII fragments relative to the A. thaliana chloroplast genome. B. napus sequences located in the IR regions aligned as expected with two different positions on the Arabidopsis plastid genome. A total of five of the HindIII clones contained a fragment mapping to the IR region.

Southern blot hybridisations, in which the probes comprised previously characterised chloroplast DNA fragments from other species, were initially used to confirm the positional mapping of several HindIII library clones. For example, probes derived from the rrn region in Solanum nigrum and the rbcL region in N. tabacum were used to identify B. napus clones containing homologous sequences (Figure 2.10). These were provisionally mapped relative to the N. tabacum /A. thaliana plastid genomes and their positions subsequently confirmed by sequence analysis.

It was not possible however to obtain a complete clone coverage of the B. napus plastid genome using only the HindIII clone library. For this reason, a SalI clone library was made by cloning SalI digested chloroplast DNA into pUC7. E. coli transformants containing chloroplast DNA fragments were selected by colony blot hybridisation using α-[P32]-dCTP labelled total chloroplast DNA as the probe. Again minipreps were screened by digestion with the cloning enzyme alone, followed by double-digestion with both SalI and EcoR1 to identify unique clones. Despite repeated cloning attempts and the large number of transformants screened, only three unique SalI clones were isolated. Figure 2.7 shows that the three SalI clones contained a single plastid DNA insert with a size greater than 10kb. To demonstrate that these clones did in fact contain chloroplast DNA fragments, a Southern blot analysis was performed. Figure 2.8 clearly shows that the total chloroplast DNA probe hybridised to all three SalI clones digested with both SalI and EcoR1. As with the HindIII clones, the ends of the SalI fragment inserts were subsequently sequenced and the sequences aligned with the N. tabacum /A. thaliana plastid genome maps. These map alignments are shown in Table 2.2.
They revealed that one of the clones (Sal 36) contained a fragment localised to the IR regions, with the other two cloned fragments being mapped to the large single copy (LSC) region.

Because of the limited success of the Sal1 library, an EcoR1 clone library was also prepared using pUC7. The smaller EcoR1 DNA fragments proved easier to clone. Following screening by both single (EcoR1) and double (EcoR1 + BamHI) digestions, a total of 21 unique clones were isolated. Figure 2.6 shows the plastid DNA inserts in 21 unique EcoR1 clones released by digestion with the EcoR1. All clones contain a single chloroplast DNA fragment that ranges in size from approximately 7.5kb to less than 1kb. Each of these were end-sequenced and the sequences were aligned with the N. tabacum / A. thaliana maps. These results are detailed in Table 2.3. In this way, five of the cloned B. napus plastid DNA fragments were localized to the IR region.

A direct comparison of the existing B. napus cloned fragments (i.e. the various HindIII, Sal1 and EcoR1 clones) with the Arabidopsis plastid genome map revealed two notable gaps in the overall clone coverage (Figure 2.3): one in the Large Single Copy region (encompassing rps16) and the other at the IRa/Small Single Copy (SSC) junction. In order to attempt to close these gaps, large XbaI plastid DNA fragments were identified by Southern hybridisation using probes derived from clones adjacent to the gaps. Unfortunately in the case of a 10 kb gap extending from rpl2 to atpA only two small EcoR1 clones (Eco 80 & Eco 73) could be found which did little to provide coverage of the region. In a final attempt to close this gap, radiolabeled probes derived from Eco 80 and Eco 73 were used in a Southern blot analysis of plastid DNA digests to identify restriction fragments of a size likely to extend across the gap region. A candidate XbaI fragment of approximately 7 kb was identified using this approach (Figure 2.9). The same radiolabeled probes were then used to screen E. coli colonies containing plastid DNA XbaI fragments cloned into pUC19 (data not shown). This identified the clone Xba 8, which contained a 4.3kb insert covering the rps16 region localized to the middle of the 10 kb gap. Colony hybridisation also identified the clone Xba 17, which contained a 15.8kb insert covering the gap in the SSC/IRa junction (data not
shown). All the XbaI clones are described in Table 2.4 and restriction digest analysis of the clones is shown in Figure 2.7.

To summarise, restriction fragment clone libraries of B. napus chloroplast DNA were constructed using four restriction endonucleases; HindIII, EcoR1, SalI and XbaI. A total of 50 unique clones were characterised in detail, averaging 5kb in size whose combined length gave approximately 97% total coverage of the chloroplast genome. Taken individually, 23 HindIII clones covered approximately 74% of the genome, but the others gave substantially less coverage; 21 EcoR1 clones covered 42%, 3 SalI clones covered 25%, and 3 XbaI clones covered 23% of the genome. Figure 2.3 shows a detailed map of all the clones aligned with the A. thaliana chloroplast gene map (Sato et al., 1999).

2.3.2. Nucleotide Sequencing of the B. napus Plastid Genome

Three large regions of the B. napus plastid genome were sequenced by means of primer walking along the mapped clone bank DNAs. The first of these is the 11kb region spanning the clones Hin 27 and Hin32, which includes the genes \( rbcl \) and \( accD \). The second is a 9.4kb region found in the inverted repeats, spanning the clones Hin 18 and Eco 31 and including the 16S rRNA gene region. Finally, sequence was also determined for a 36.1kb region that contains the genes \( rpoB \), \( rpoC1 \) and \( rpoC2 \). Maps of these sequenced regions, including the genes identified and their relationship to the cloned fragments are shown in Figures 2.11 and 2.12.

Having assembled the nucleotide sequence of the three regions described, they were analysed as follows. Firstly, the open reading frames (orfs) were identified by comparison with the Arabidopsis plastid genome nucleotide sequence (Sato et al., 1999; Ac. AP000423). The possibility of the B. napus sequence containing additional genes was ruled out on the basis of a BLASTP search against the non-redundant protein database (Altschul et al., 1990). Analysis of the combined 56.6kb nucleotide sequence showed that it codes for a total of 29 polypeptides (26 complete and another 3 partial orfs \( rps12, ndhB \) and \( cemA \)), 2 rRNA species (complete 16S rRNA gene; 23S rRNA gene in part) and 15 tRNAs.
2.3.2.1. Phylogenetic Analysis.

The sequences of all of the protein- and rRNA-coding genes (tRNA genes excluded) identified in the *B. napus* sequence data were individually aligned with the corresponding homologues from those higher plant species whose entire plastid genome nucleotide sequences are available in GenBank. These species include four other dicots *A. thaliana* (Sato et al., 1999), *N. tabacum* (Shinozaki et al., 1986), *Spinacia oleracea* (Schmitz-Linneweber et al., 2001), and *Oenothera elata* (Hupfer et al., 2000), two monocots *Oryza sativa* (Hiratsuka et al., 1989), *Zea mays* (Maier et al., 1995), and a gymnosperm *Pinus thunbergii* (Wakasugi et al., 1994). Sequence similarities were scored by percentage identity. This was performed both at the nucleotide and amino acid levels. Table 2.5 shows the percentage nucleotide base identities for the 31 selected genes. Unsurprisingly, the ribosomal RNA genes (*rrn16S* and *rrn23S*) displayed extremely high conservation across all the species analysed. Some protein-coding genes, such as *psaB* and *rps7*, were highly conserved. Other genes, such as *psal*, *rpoC2* and *rps14* possessed lower than average percentage base identities. The nucleotide sequence of the *accD* gene was particularly divergent across the species, with no functional homologue present in the plastid genomes of rice (Hiratsuka et al., 1989) or maize (Maier et al., 1995).

Table 2.6 contains the percentage identity values for the translated amino acid sequences of the 29 protein-coding genes selected. A total 26 of these had functional homologues common to all 8 species. The excluded orfs were the *ndhB* and *ndhC* genes, which are absent from the plastid genome of pine (Wakasugi et al., 1994) and also the *accD* gene. When the 26 orfs were joined end to end, the concatenated sequence codes for a total of 8923 amino acids in *B. napus*. The homologous orfs were also joined in the same order for the other species. The concatenated amino acid sequences representing the seven selected angiosperm species (pine was excluded as it is a gymnosperm) were used in a CLUSTALW multi-sequence alignment (Thompson et al., 1994). Protein sequences were used because of the inconsistent frequency of A + T-rich codons across the plastid genomes (Martin et al., 1998). This alignment was subsequently used to construct various neighbor-joining trees (Saitou and Nei, 1987), which provided visual representation of the phylogenetic relationships existing between the flowering
plant species. Figure 2.13 contains a phylogenetic tree in the phenogram style. Bootstrap values based on 1000 replicates were calculated using PAUP* version 4.0 (Swofford, 1998). Figure 2.14 contains a tree that clearly highlights the relatively recent common ancestry that exists between the *B. napus* and *A. thaliana* species.

Assuming amino acid substitution rates to be constant (a criteria unlikely to be entirely true), the dimensions of the trees are proportionally representative of evolutionary distance. Using an internal calibration point, placing the monocot-dicot divergence date at 160-240 million years (Wolfe *et al.*, 1989), divergence dates for all nodes on the tree were estimated. Tables 2.7 details the individual branch lengths, measured in units relative to the distance from nodes 1 to 5 (Figure 2.14). The problem of uneven branch lengths (indicative of unequal substitution rates) interfering with date calculations was reduced by the use of mean lengths. Table 2.8 contains a list of the estimated angiosperm divergence dates, which places the *Brassica-Arabidopsis* split at between 14.46 and 21.69 Myr. A nucleotide substitution rate of 0.507-0.76 x 10^{-9} substitutions per site per year for plastid coding sequence was calculated from the average *Brassica-Arabidopsis* coding sequence identity value (97.8%) using Kimura’s two-parameter method (Kimura, 1980). The details of this calculation are outlined in Table 2.10.

### 2.3.2.2. Intergenic Sequence Analysis.

Sequence analysis of two intergenic regions was also performed because of their relevance for plastid transformation vector construction (see Chapter 3). Both these non-coding sequences, namely the 1919 bp nucleotide sequence between the coding regions of *trnV* and *rps12 3'* in the inverted repeat and the 674 bp sequence located between *rbcL* and *accD*, were aligned with the same seven plastid species sequences used for the phylogenetic analysis of coding regions. The nucleotide sequence alignments shown in Figures 2.15 and 2.16 clearly demonstrate the accumulation of numerous substitutions, insertions and deletions in both these regions, even between species as closely related as *A. thaliana* and *B. napus*. In fact, in the case of the *rbcL–accD* spacer region, the mutation rate is so great that no significant alignment can be detected between *B. napus* and any other species more distantly related than *A. thaliana*. 58
The entire \textit{rbcl}, \textit{accD} and spacer region of \textit{B. napus} was subdivided into short sequence sections (between 135 and 300 bp) and aligned with the corresponding region of the plastid genomes of both tobacco (Shinozaki \textit{et al}, 1986) and \textit{Arabidopsis} (Sato \textit{et al}, 1999) with the aim of identifying any conserved/non-conserved domains (Figure 2.17). This analysis revealed that the 5' end (approximately the first 600 bp) of the \textit{accD} gene is considerably less conserved between all 3 species than is the 3' end of the gene. Similarly the 3' 300 bp of the \textit{rbcl} gene is also less conserved between \textit{B. napus} and \textit{N. tabacum} genomes than is the 5' end of the gene.

Unlike the \textit{rbcl-accD} intergenic sequence alignments, sensible alignments were obtained for the \textit{Brassica trnV-rps12} 3' spacer region with the homologous sequence in all of the other selected species, including that of pine. Figure 2.15 provides sampled regions of three such alignments: those of the \textit{B. napus} sequence aligned to the homologous regions of the \textit{A. thaliana}, \textit{N. tabacum} and \textit{Z. mays} plastid genomes. In keeping with the phylogenetic conclusions drawn from the protein-coding sequence analysis, it was clear that the \textit{Brassica trnV-rps12} 3' intergenic sequence was less divergent from the \textit{Arabidopsis} sequence than it was from the tobacco or maize homologous sequences.

The percentage nucleotide identity values were estimated for all of the intergenic sequence alignments described above. These values are contained in Table 2.9, which provides a direct comparison with the average plastid coding sequence identity values. The \textit{trnV-rps12} 3' intergenic sequence identity values were clearly reduced compared to the coding sequence values for all of the interspecies alignments. The percentage identity of the 674 bp \textit{Brassica rbcl-accD} spacer sequence with that of \textit{Arabidopsis} was considerably reduced (80.4%) compared to the average coding sequence identity between the same two species (97.8%). The lack of sensible alignments between the \textit{Brassica rbcl-accD} intergenic sequence and the corresponding region in any of the other species meant that no identity values could be attributed to this region.
The nucleotide substitution rates of both the non-coding regions described were independently calculated from their respective *Brassica-Arabidopsis* sequence alignments. These calculations are detailed in Table 2.10. The *trnV–rps12* 3' spacer has a nucleotide substitution rate of $1.452-2.179 \times 10^{-9}$ substitutions per site per year, approximately 3 times greater than the average plastid coding sequence rate (calculated previously to be $0.507-0.76 \times 10^{-9}$ substitutions per site per year). The substitution rate for the *rbcl–accD* spacer sequence is $4.518-6.778 \times 10^{-9}$ substitutions per site per year, approximately 9 times greater than the genic rate.
Figure 2.1. Commercial Value of Oilseed Rape. Pictures A and B show the flowering plant but it is the seed which holds the commercial value. As well as providing high grade animal feed its oil is becoming increasingly popular in cooking.

Figure 2.2. Disease and Pest Management in Oilseed Rape. Picture E is of the Bertha armyworm, which feeds on the leaves of *B. napus* and can reduce yields. Picture F demonstrates the deleterious effect an infection like brown girdling root rot has on the health and subsequently the yields of the plant. Pictures G and H demonstrate methods used to combat such problems, the spraying of chemicals and most recently the use of genetically modified plants.
Figure 2.3. *Brassica napus* Chloroplast Clone Bank. The outer circle represents the *A. thaliana* plastome. Inside the circle, individual clones from the four *B. napus* plastome libraries have been aligned with the *A. thaliana* gene map giving a visual description of clone coverage. This figure was generated using the clone insert end sequences. Hence, the positioning of the individual clones relative to the *A. thaliana* plastid genome is accurate and to scale (Vector NTI software). The transgene integration position for both of the 5ras5/ca-specific transformation vectors pPCl and pZBl (Chapter 3) are also illustrated. Note: not all chloroplast genes are labeled.
Figure 2.4. Restriction Analysis of Purified Chloroplast DNA from *B. napus*. The enzymes *EcoR1* and *BamH1* cut the chloroplast genome frequently whereas the *Sal1* and *Kpn1* digests give fewer but bigger fragments.

Figure 2.5. Cloned *HindIII* *B. napus* Chloroplast DNA Fragments. Inserts have been cut from the 2.6kb vector pIC19H by *HindIII* digestion. Note: the clone *Hin126* has incorporated 2 different fragments.

Figure 2.6. Cloned *EcoR1* *B. napus* Chloroplast DNA Fragments. Inserts have been cut from the 2.6kb vector pUC7 by *EcoR1* digestion.
Figure 2.7. *SalI* and *XbaI* *B. napus* Chloroplast DNA Restriction Fragment Clones. The left-hand side of the electrophoresed agarose gel shows the three *SalI* clones cut from the 2.6kb vector pUC7 by *SalI* digestion. To the right-hand side of the gel are the *XbaI* clones cut from pUC19 by digestion with the cloning enzyme *XbaI*.

Figure 2.8. Southern Blot Analysis of *B. napus* Clones. **Picture A** shows the *SalI* clones digested with both *SalI* and *EcoR1*, electrophoresed on 0.8% agarose and stained with Ethidium Bromide. **Picture B** shows an autoradiograph of the same gel, transferred to nitrocellulose and probed with α-[P32]-dCTP labelled chloroplast DNA cut with *SalI*. Note: the 2.6kb vector band is clearly absent from the autoradiograph. All other bands are clearly chloroplast derived.
Figure 2.9. Southern Blot Analysis of Chloroplast DNA Digests. Four chloroplast DNA digests were electrophoresed and transferred to nitrocellulose. **Autoradiograph A** was probed with the gel purified chloroplast DNA fragment from the clone Eco 73, labelled with α-[P32]-dCTP. Note: as control the purified insert for this clone was also run on the gel. **Autoradiograph B** was probed with the gel purified insert from the clone Eco 80, labelled with α-[P32]-dCTP.

Figure 2.10. Cross Hybridisation of *B. napus* Clones with Chloroplast Sequences from other Species. A selection of the *B. napus* chloroplast *HindIII* clone inserts, electrophoresed and transferred to nitrocellulose. Autoradiograph A has been probed with the α-[P32]-dCTP labelled *S. nigrum* chloroplast transformation vector pSSH11 (Kavanagh *et al.*, 1999). Hybridisation has occurred with clones containing the 16S rDNA region only. Autoradiograph B has been probed with the α-[P32]-dCTP labelled *N. tabacum* chloroplast transformation vector pZS197 (Svab and Maliga, 1993). This probe only hybridised to those clones containing chloroplast DNA from the *rbcL* region.
Table 2.1. *HindIII* Clones of *B. napus* Chloroplast DNA. Arranged in order of size. The fragments were mapped by aligning either end sequence of the insert with the complete chloroplast nucleotide sequence of *A. thaliana* (Ac. AP000423). Those end sequences located in the IR regions aligned with two different positions on the *Arabidopsis* plastome.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size (approx.)</th>
<th><em>A. thaliana</em> match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hin 50</td>
<td>12.3kb</td>
<td>22292-34642</td>
</tr>
<tr>
<td>Hin 123</td>
<td>11.6kb</td>
<td>84823-96433</td>
</tr>
<tr>
<td></td>
<td></td>
<td>142212-153822</td>
</tr>
<tr>
<td>Hin 120</td>
<td>8.9kb</td>
<td>34643-43553</td>
</tr>
<tr>
<td>Hin 13</td>
<td>8.7kb</td>
<td>113538-122196</td>
</tr>
<tr>
<td>Hin 16</td>
<td>7.6kb</td>
<td>61190-68795</td>
</tr>
<tr>
<td>Hin 144</td>
<td>7.2kb</td>
<td>76980-84218</td>
</tr>
<tr>
<td>Hin 27</td>
<td>5.8kb</td>
<td>50073-55904</td>
</tr>
<tr>
<td>Hin 43</td>
<td>5.7kb</td>
<td>122197-127882</td>
</tr>
<tr>
<td>Hin 133</td>
<td>5.5kb</td>
<td>108077-113537</td>
</tr>
<tr>
<td>Hin 32</td>
<td>5.2kb</td>
<td>55905-61128</td>
</tr>
<tr>
<td>Hin 18</td>
<td>5.2kb</td>
<td>96434-101607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>137038-142211</td>
</tr>
<tr>
<td>Hin 86</td>
<td>4.7kb</td>
<td>15553-20268</td>
</tr>
<tr>
<td>Hin 34</td>
<td>3.8kb</td>
<td>101608-105406</td>
</tr>
<tr>
<td></td>
<td></td>
<td>133239-137037</td>
</tr>
<tr>
<td>Hin 107</td>
<td>3.7kb</td>
<td>121867-127882</td>
</tr>
<tr>
<td>Hin 139</td>
<td>3.1kb</td>
<td>46341-49477</td>
</tr>
<tr>
<td>Hin 28</td>
<td>2.8kb</td>
<td>70432-73269</td>
</tr>
<tr>
<td>Hin 126</td>
<td>2.7kb (gel</td>
<td>43605-?</td>
</tr>
<tr>
<td></td>
<td>estimate)</td>
<td>122197-?</td>
</tr>
<tr>
<td></td>
<td>+ 5.7kb (Hin 43)</td>
<td></td>
</tr>
<tr>
<td>Hin 141</td>
<td>2.2kb</td>
<td>13322-15552</td>
</tr>
<tr>
<td>Hin 100</td>
<td>2.0kb</td>
<td>20269-22291</td>
</tr>
<tr>
<td>Hin 66</td>
<td>1.7kb</td>
<td>11640-13321</td>
</tr>
<tr>
<td>Hin 60</td>
<td>1.4kb</td>
<td>68796-70153</td>
</tr>
<tr>
<td>Hin 103</td>
<td>1.3kb</td>
<td>106776-108076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>130569-131869</td>
</tr>
<tr>
<td>Hin 82</td>
<td>0.8kb</td>
<td>10853-11639</td>
</tr>
</tbody>
</table>

Table 2.2. *SalI* Clones of *B. napus* Chloroplast DNA.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size (approx.)</th>
<th><em>A. thaliana</em> match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sal 25</td>
<td>16.3kb</td>
<td>68181-84481</td>
</tr>
<tr>
<td>Sal 36</td>
<td>11.6kb</td>
<td>84482-96052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>142593-154163</td>
</tr>
<tr>
<td>Sal 20</td>
<td>10.1kb</td>
<td>9373-19453</td>
</tr>
</tbody>
</table>
Table 2.3. **EcoRI Clones of *B. napus* Chloroplast DNA.**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size (approx.)</th>
<th>A. thaliana match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco 115</td>
<td>7.6kb</td>
<td>28387-35938</td>
</tr>
<tr>
<td>Eco 5</td>
<td>7.5kb</td>
<td>36250-43778</td>
</tr>
<tr>
<td>Eco 24</td>
<td>6.7kb</td>
<td>110941-117677</td>
</tr>
<tr>
<td>Eco 73</td>
<td>5.0kb</td>
<td>9509-14531</td>
</tr>
<tr>
<td>Eco 29</td>
<td>4.7kb</td>
<td>49928-54600</td>
</tr>
<tr>
<td>Eco 31</td>
<td>4.6kb</td>
<td>102334-106896, 131749-136311</td>
</tr>
<tr>
<td>Eco 26</td>
<td>4.4kb</td>
<td>66170-70553</td>
</tr>
<tr>
<td>Eco 27</td>
<td>3.8kb</td>
<td>22202-26004</td>
</tr>
<tr>
<td>Eco 63</td>
<td>2.4kb</td>
<td>99981-102333, 136312-138664</td>
</tr>
<tr>
<td>Eco 4</td>
<td>2.2kb</td>
<td>44193-46375</td>
</tr>
<tr>
<td>Eco 37</td>
<td>2.1kb</td>
<td>93974-96122, 142523-144671</td>
</tr>
<tr>
<td>Eco 7</td>
<td>2.0kb</td>
<td>91068-93044, 145600-147577</td>
</tr>
<tr>
<td>Eco 80</td>
<td>1.9kb</td>
<td>2185-4108</td>
</tr>
<tr>
<td>Eco 1</td>
<td>1.9kb</td>
<td>74563-76456</td>
</tr>
<tr>
<td>Eco 53</td>
<td>1.8kb</td>
<td>72790-74562</td>
</tr>
<tr>
<td>Eco 100</td>
<td>1.6kb</td>
<td>61159-62767</td>
</tr>
<tr>
<td>Eco 12</td>
<td>1.5kb</td>
<td>63934-65473</td>
</tr>
<tr>
<td>Eco 77</td>
<td>1.5kb</td>
<td>98136-99615, 139030-140509</td>
</tr>
<tr>
<td>Eco 70</td>
<td>0.9kb</td>
<td>82141-83002</td>
</tr>
<tr>
<td>Eco 8</td>
<td>0.8kb</td>
<td>153748-00035</td>
</tr>
<tr>
<td>Eco 94</td>
<td>0.7kb</td>
<td>46834-47545</td>
</tr>
</tbody>
</table>

Table 2.4. **XbaI Clones of *B. napus* Chloroplast DNA.**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size (approx.)</th>
<th>A. thaliana match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xba 17</td>
<td>15.8kb</td>
<td>123096-138943</td>
</tr>
<tr>
<td>Xba 16</td>
<td>15.4kb</td>
<td>99702-115103</td>
</tr>
<tr>
<td>Xba 8</td>
<td>4.3kb</td>
<td>3087-7415</td>
</tr>
</tbody>
</table>
Figure 2.11. Maps of the two regions used for the construction of the *Brassica* plastid transformation vectors “pZB1” and “pPC1”. The upper levels show the genetic map of the sequenced region and beneath are the relevant clones. Colour codes are the same as for Figure 2.3.
Figure 2.12. Map of the 36kb region surrounding the rpoB gene in the B. napus plastid genome. The upper level shows the genetic map of the sequenced region and beneath are the relevant cloned fragments.
<table>
<thead>
<tr>
<th></th>
<th>A. thaliana</th>
<th>N. tabacum</th>
<th>S. oleracea</th>
<th>O. elata</th>
<th>Z. mays</th>
<th>O. sativa</th>
<th>P. thunbergii</th>
</tr>
</thead>
<tbody>
<tr>
<td>rps2</td>
<td>98.3%</td>
<td>90.0%</td>
<td>88.2%</td>
<td>87.1%</td>
<td>81.2%</td>
<td>81.9%</td>
<td>73.0%</td>
</tr>
<tr>
<td>rps4</td>
<td>97.5%</td>
<td>89.6%</td>
<td>86.8%</td>
<td>87.8%</td>
<td>81.7%</td>
<td>81.8%</td>
<td>62.9%</td>
</tr>
<tr>
<td>rps7</td>
<td>99.4%</td>
<td>97.9%</td>
<td>98.1%</td>
<td>95.5%</td>
<td>91.5%</td>
<td>91.5%</td>
<td>89.7%</td>
</tr>
<tr>
<td>rps12 3'</td>
<td>99.6%</td>
<td>98.4%</td>
<td>99.2%</td>
<td>98.1%</td>
<td>95.0%</td>
<td>95.0%</td>
<td>95.7%</td>
</tr>
<tr>
<td>rps14</td>
<td>97.7%</td>
<td>89.0%</td>
<td>85.3%</td>
<td>86.7%</td>
<td>77.7%</td>
<td>73.7%</td>
<td>52.3%</td>
</tr>
<tr>
<td>petN</td>
<td>98.9%</td>
<td>92.2%</td>
<td>90.0%</td>
<td>87.7%</td>
<td>87.7%</td>
<td>88.9%</td>
<td>80.0%</td>
</tr>
<tr>
<td>psaA</td>
<td>94.3%</td>
<td>92.5%</td>
<td>91.7%</td>
<td>91.7%</td>
<td>89.7%</td>
<td>89.4%</td>
<td>85.0%</td>
</tr>
<tr>
<td>psaB</td>
<td>98.6%</td>
<td>92.9%</td>
<td>92.7%</td>
<td>92.3%</td>
<td>90.4%</td>
<td>90.5%</td>
<td>90.2%</td>
</tr>
<tr>
<td>psaI</td>
<td>98.2%</td>
<td>75.4%</td>
<td>81.6%</td>
<td>60.5%</td>
<td>57.0%</td>
<td>59.6%</td>
<td>40.4%</td>
</tr>
<tr>
<td>psbC</td>
<td>98.4%</td>
<td>92.6%</td>
<td>91.4%</td>
<td>93.1%</td>
<td>89.5%</td>
<td>89.5%</td>
<td>87.5%</td>
</tr>
<tr>
<td>psbD</td>
<td>98.8%</td>
<td>94.5%</td>
<td>94.2%</td>
<td>94.0%</td>
<td>89.1%</td>
<td>90.7%</td>
<td>88.9%</td>
</tr>
<tr>
<td>psbM</td>
<td>97.1%</td>
<td>88.6%</td>
<td>90.5%</td>
<td>82.9%</td>
<td>86.7%</td>
<td>88.6%</td>
<td>66.7%</td>
</tr>
<tr>
<td>atpA</td>
<td>97.6%</td>
<td>88.3%</td>
<td>74.1%</td>
<td>87.5%</td>
<td>82.3%</td>
<td>82.6%</td>
<td>76.2%</td>
</tr>
<tr>
<td>atpB</td>
<td>96.7%</td>
<td>88.8%</td>
<td>87.6%</td>
<td>87.7%</td>
<td>84.2%</td>
<td>81.0%</td>
<td>78.9%</td>
</tr>
<tr>
<td>atpE</td>
<td>97.0%</td>
<td>88.2%</td>
<td>90.0%</td>
<td>88.2%</td>
<td>79.4%</td>
<td>76.7%</td>
<td>75.7%</td>
</tr>
<tr>
<td>atpF</td>
<td>95.3%</td>
<td>86.7%</td>
<td>87.2%</td>
<td>73.5%</td>
<td>77.3%</td>
<td>76.9%</td>
<td>67.2%</td>
</tr>
<tr>
<td>atpH</td>
<td>98.8%</td>
<td>91.5%</td>
<td>97.4%</td>
<td>90.7%</td>
<td>89.0%</td>
<td>89.0%</td>
<td>83.3%</td>
</tr>
<tr>
<td>atpI</td>
<td>97.1%</td>
<td>88.5%</td>
<td>83.6%</td>
<td>88.5%</td>
<td>80.8%</td>
<td>81.3%</td>
<td>79.4%</td>
</tr>
<tr>
<td>ndhB</td>
<td>100%</td>
<td>97.9%</td>
<td>97.4%</td>
<td>95.6%</td>
<td>96.1%</td>
<td>95.8%</td>
<td></td>
</tr>
<tr>
<td>ndhC</td>
<td>98.1%</td>
<td>88.7%</td>
<td>90.6%</td>
<td>90.1%</td>
<td>84.8%</td>
<td>84.8%</td>
<td></td>
</tr>
<tr>
<td>rbcL</td>
<td>97.5%</td>
<td>88.6%</td>
<td>88.5%</td>
<td>88.7%</td>
<td>84.1%</td>
<td>84.2%</td>
<td>84.1%</td>
</tr>
<tr>
<td>rpoB</td>
<td>97.1%</td>
<td>89.2%</td>
<td>88.4%</td>
<td>89.3%</td>
<td>80.9%</td>
<td>80.4%</td>
<td>74.6%</td>
</tr>
<tr>
<td>rpoCl</td>
<td>98.0%</td>
<td>90.5%</td>
<td>89.7%</td>
<td>87.3%</td>
<td>77.0%</td>
<td>77.4%</td>
<td>62.2%</td>
</tr>
<tr>
<td>rpoC2</td>
<td>94.5%</td>
<td>79.7%</td>
<td>73.8%</td>
<td>78.5%</td>
<td>56.8%</td>
<td>55.1%</td>
<td>39.6%</td>
</tr>
<tr>
<td>accD</td>
<td>95.3%</td>
<td>65.2%</td>
<td>64.7%</td>
<td>38.6%</td>
<td></td>
<td></td>
<td>33.3%</td>
</tr>
<tr>
<td>cemA</td>
<td>97.7%</td>
<td>82.7%</td>
<td>77.9%</td>
<td>68.2%</td>
<td>65.9%</td>
<td>60.1%</td>
<td>31.3%</td>
</tr>
<tr>
<td>ycf3</td>
<td>99.2%</td>
<td>92.7%</td>
<td>92.7%</td>
<td>91.1%</td>
<td>87.9%</td>
<td>87.9%</td>
<td>80.3%</td>
</tr>
<tr>
<td>ycf4</td>
<td>96.6%</td>
<td>86.5%</td>
<td>86.3%</td>
<td>83.6%</td>
<td>82.7%</td>
<td>82.0%</td>
<td>72.6%</td>
</tr>
<tr>
<td>ycf9</td>
<td>98.4%</td>
<td>92.6%</td>
<td>90.5%</td>
<td>91.0%</td>
<td>84.1%</td>
<td>84.1%</td>
<td>81.0%</td>
</tr>
<tr>
<td>rrn16S</td>
<td>100%</td>
<td>99.5%</td>
<td>99.1%</td>
<td>98.2%</td>
<td>96.5%</td>
<td>96.6%</td>
<td>97.1%</td>
</tr>
<tr>
<td>rrn23S</td>
<td>99.9%</td>
<td>98.2%</td>
<td>97.1%</td>
<td>97.6%</td>
<td>95.7%</td>
<td>96.4%</td>
<td>95.2%</td>
</tr>
<tr>
<td>Average</td>
<td>97.79%</td>
<td>89.60%</td>
<td>88.41%</td>
<td>86.17%</td>
<td>83.42%</td>
<td>83.11%</td>
<td>73.25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ribosomal protein genes</th>
<th>Plastid RNA polymerase and other genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthesis genes</td>
<td></td>
</tr>
<tr>
<td>rRNA genes</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5. *B. napus* Chloroplast Nucleotide Sequence Similarities. Percentage nucleotide identity of the 31 *Brassica* plastid coding sequences with those of the seven other species obtained from GenBank.
<table>
<thead>
<tr>
<th></th>
<th>A. thaliana</th>
<th>N. tabacum</th>
<th>S. oleracea</th>
<th>O. elata</th>
<th>Z. mays</th>
<th>O. sativa</th>
<th>P. thunbergii</th>
</tr>
</thead>
<tbody>
<tr>
<td>rps2</td>
<td>99.6%</td>
<td>89.4%</td>
<td>89.4%</td>
<td>83.1%</td>
<td>76.3%</td>
<td>77.1%</td>
<td>69.9%</td>
</tr>
<tr>
<td>rps4</td>
<td>98.0%</td>
<td>90.0%</td>
<td>86.1%</td>
<td>86.1%</td>
<td>79.6%</td>
<td>80.6%</td>
<td>63.7%</td>
</tr>
<tr>
<td>rps7</td>
<td>100%</td>
<td>97.4%</td>
<td>96.8%</td>
<td>87.7%</td>
<td>84.5%</td>
<td>83.9%</td>
<td>83.9%</td>
</tr>
<tr>
<td>rps12 3'</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>96.5%</td>
<td>92.9%</td>
<td>92.9%</td>
<td>95.3%</td>
</tr>
<tr>
<td>rps14</td>
<td>97.0%</td>
<td>88.9%</td>
<td>82.8%</td>
<td>82.8%</td>
<td>86.9%</td>
<td>85.9%</td>
<td>76.8%</td>
</tr>
<tr>
<td>petN</td>
<td>96.6%</td>
<td>96.6%</td>
<td>96.6%</td>
<td>96.6%</td>
<td>93.1%</td>
<td>96.6%</td>
<td>86.2%</td>
</tr>
<tr>
<td>psaA</td>
<td>99.2%</td>
<td>98.0%</td>
<td>97.5%</td>
<td>96.7%</td>
<td>96.4%</td>
<td>95.5%</td>
<td>92.3%</td>
</tr>
<tr>
<td>psaB</td>
<td>99.6%</td>
<td>97.4%</td>
<td>97.7%</td>
<td>96.6%</td>
<td>96.7%</td>
<td>96.7%</td>
<td>92.1%</td>
</tr>
<tr>
<td>psaI</td>
<td>97.3%</td>
<td>78.4%</td>
<td>73.0%</td>
<td>64.9%</td>
<td>81.1%</td>
<td>75.7%</td>
<td>48.6%</td>
</tr>
<tr>
<td>psbC</td>
<td>100%</td>
<td>98.5%</td>
<td>97.9%</td>
<td>97.0%</td>
<td>95.1%</td>
<td>96.4%</td>
<td>95.3%</td>
</tr>
<tr>
<td>psbD</td>
<td>100%</td>
<td>98.9%</td>
<td>99.4%</td>
<td>98.9%</td>
<td>94.9%</td>
<td>98.0%</td>
<td>96.6%</td>
</tr>
<tr>
<td>psbM</td>
<td>97.1%</td>
<td>97.1%</td>
<td>97.1%</td>
<td>94.1%</td>
<td>94.1%</td>
<td>97.1%</td>
<td>79.4%</td>
</tr>
<tr>
<td>atpA</td>
<td>98.6%</td>
<td>93.1%</td>
<td>78.1%</td>
<td>90.1%</td>
<td>85.2%</td>
<td>85.6%</td>
<td>84.0%</td>
</tr>
<tr>
<td>atpB</td>
<td>97.8%</td>
<td>92.4%</td>
<td>94.0%</td>
<td>90.8%</td>
<td>90.0%</td>
<td>90.8%</td>
<td>87.1%</td>
</tr>
<tr>
<td>atpE</td>
<td>98.5%</td>
<td>88.6%</td>
<td>91.7%</td>
<td>92.4%</td>
<td>76.5%</td>
<td>74.2%</td>
<td>71.2%</td>
</tr>
<tr>
<td>atpF</td>
<td>95.7%</td>
<td>87.0%</td>
<td>86.4%</td>
<td>76.6%</td>
<td>71.2%</td>
<td>72.8%</td>
<td>63.6%</td>
</tr>
<tr>
<td>atpH</td>
<td>100%</td>
<td>98.8%</td>
<td>97.5%</td>
<td>97.5%</td>
<td>97.5%</td>
<td>97.5%</td>
<td>96.3%</td>
</tr>
<tr>
<td>atpI</td>
<td>98.0%</td>
<td>94.0%</td>
<td>92.4%</td>
<td>90.8%</td>
<td>90.8%</td>
<td>91.2%</td>
<td>86.7%</td>
</tr>
<tr>
<td>ndhB</td>
<td>100%</td>
<td>99.6%</td>
<td>98.3%</td>
<td>94.4%</td>
<td>96.5%</td>
<td>96.5%</td>
<td>96.5%</td>
</tr>
<tr>
<td>ndhC</td>
<td>98.3%</td>
<td>90.8%</td>
<td>90.8%</td>
<td>87.5%</td>
<td>85.8%</td>
<td>85.8%</td>
<td>85.8%</td>
</tr>
<tr>
<td>rbcL</td>
<td>98.7%</td>
<td>92.9%</td>
<td>91.6%</td>
<td>93.7%</td>
<td>91.0%</td>
<td>92.7%</td>
<td>92.5%</td>
</tr>
<tr>
<td>rpoB</td>
<td>99.3%</td>
<td>91.5%</td>
<td>91.1%</td>
<td>89.9%</td>
<td>79.7%</td>
<td>80.1%</td>
<td>71.1%</td>
</tr>
<tr>
<td>rpoC1</td>
<td>98.2%</td>
<td>91.9%</td>
<td>91.6%</td>
<td>86.5%</td>
<td>78.7%</td>
<td>78.3%</td>
<td>68.3%</td>
</tr>
<tr>
<td>rpoC2</td>
<td>94.9%</td>
<td>80.7%</td>
<td>78.9%</td>
<td>78.3%</td>
<td>62.8%</td>
<td>67.8%</td>
<td>49.3%</td>
</tr>
<tr>
<td>accD</td>
<td>91.6%</td>
<td>66.3%</td>
<td>64.4%</td>
<td>38.0%</td>
<td>38.0%</td>
<td>38.0%</td>
<td>38.0%</td>
</tr>
<tr>
<td>cemA</td>
<td>97.7%</td>
<td>73.8%</td>
<td>70.0%</td>
<td>62.3%</td>
<td>63.1%</td>
<td>59.2%</td>
<td>40.8%</td>
</tr>
<tr>
<td>ycf3</td>
<td>100%</td>
<td>96.0%</td>
<td>94.4%</td>
<td>95.2%</td>
<td>92.1%</td>
<td>92.1%</td>
<td>85.7%</td>
</tr>
<tr>
<td>ycf4</td>
<td>95.7%</td>
<td>85.3%</td>
<td>83.7%</td>
<td>83.2%</td>
<td>81.0%</td>
<td>80.4%</td>
<td>72.3%</td>
</tr>
<tr>
<td>ycf9</td>
<td>100%</td>
<td>95.2%</td>
<td>95.2%</td>
<td>95.2%</td>
<td>83.9%</td>
<td>90.3%</td>
<td>75.8%</td>
</tr>
<tr>
<td>Average</td>
<td>98.1%</td>
<td>91.33%</td>
<td>89.81%</td>
<td>87.01%</td>
<td>85.62%</td>
<td>86.13%</td>
<td>76.43%</td>
</tr>
</tbody>
</table>

Table 2.6. B. napus Chloroplast-Encoded Protein Sequence Similarities. Percentage amino-acid identity of the 29 translated *Brassica* plastid genes sequenced with their homologues in the seven other species obtained from GenBank.
Figure 2.13. Phylogenetic Relationships within the Angiosperm Lineage derived from 26 Chloroplast Protein Sequences. The sequences of 26 selected proteins were joined end (N-terminus) to end (C-terminus) giving one large concatenated amino acid sequence. Concatenated sequences representing each of the seven selected angiosperm species were aligned using the CLUSTALW method (Thompson, et al., 1994), which in turn was used to construct this phenogram using the neighbor-joining method (Saitou and Nei, 1987), with Kimura’s method of distance measure (Kimura, 1980). Values above the branches are bootstrap percentages (Felsenstein, 1985).
Figure 2.14. Phylogenetic Tree of the Angiosperms derived from the Alignment of the concatenated Chloroplast Amino Acid Sequences. Tree constructed using the neighbor-joining method (Saitou and Nei, 1987). The branch lengths were considered to be proportional to time elapsed since divergence dates. Branch nodes are numbered 1 to 6 and divergence dates calculated in Tables 2.7 and 2.8.
Divergence | Individual Branch Lengths (Relative to Distance between Nodes 1 and 5) | Average (Units)
--- | --- | ---
Monocots-Dicots | 0.225 (Average of Rice-Maize branch distance to node 5) + 1 Unit (distance from nodes 5 to 1) | 1.225
Rosids-Asterids | 0.6196 (Average Primrose-Brassicaceae branch distance); 0.6929 (nodes 2 to 3 + {Average Spinach-Tobacco branch distance}) | 0.65625
Spinach-Tobacco | 0.8571 (Tobacco to node 3); 0.4429 (Spinach to node 3) | 0.65
Primrose-Brassicaceae | 0.6714 (Primrose to node 4); 0.5678 (nodes 6 to 4 + {Average of *B. napus* to node 6 and *A. thaliana* to node 6}) | 0.6196
Rice-Maize | 0.2215 (Rice to node 5); 0.2285 (Maize to node 5) | 0.225
*Brassica*-Arabidopsis | 0.1 (*B. napus* to node 6); 0.1214 (*A. thaliana* to node 6) | 0.1107

*Table 2.7. Angiosperm Tree Measurements.* All the branch lengths in Figure 2.14 were measured relative to the distance between nodes 1 and 5 (1 Unit). Evolutionary distances were calculated from the average branch lengths.

<table>
<thead>
<tr>
<th>Branch Node Number</th>
<th>Divergence</th>
<th>Estimated Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monocots-Dicots</td>
<td>160-240 Myr</td>
</tr>
<tr>
<td>2</td>
<td>Rosids-Asterids</td>
<td>85.71-128.57 Myr</td>
</tr>
<tr>
<td>3</td>
<td>Spinach-Tobacco</td>
<td>84.90-127.35 Myr</td>
</tr>
<tr>
<td>4</td>
<td>Primrose-Brassicaceae</td>
<td>80.93-121.40 Myr</td>
</tr>
<tr>
<td>5</td>
<td>Rice-Maize</td>
<td>29.39-44.08 Myr</td>
</tr>
<tr>
<td>6</td>
<td><em>Brassica</em>-Arabidopsis</td>
<td>14.46-21.69 Myr</td>
</tr>
</tbody>
</table>

*Table 2.8. Angiosperm tree calibration.* The calculation of Wolfe et al. (1989) placing the monocot-dicot divergence at 160-240 Myr (blue text) was used as an absolute calibration point for this tree. The divergence dates of the other five nodes were estimated using the proportional dimensions of the tree, described in Table 2.7.
Figure 2.15. Nucleotide Alignments of the Spacer Region between the *trnV* and *rps12 3' Genes in the *B. napus* Plastid Genome with the corresponding Regions in *A. thaliana* (A), *N. tabacum* (B) and *Z. mays* (C). 160 bp of the *Brassica* intergenic region was selected and aligned with the corresponding regions for *Arabidopsis*, Tobacco and Maize. The selection includes the XbaI restriction site, which was chosen as the heterologous gene-cloning site for the transformation vector pPC1. Deviations from the *Brassica* sequence, including substitutions, insertions and deletions are highlighted in yellow. Numbers in bold represent the nucleotide numbering allocated to the complete plastid sequence of respective species in the GenBank database.
Table 2.9. Intergenic Region Sequence Similarities. Percentage nucleotide identity of the *B. napus* plastid intergenic sequences with their homologues in the seven other species obtained from GenBank. The *B. napus* *trnV-rps12* 3’ spacer yielded at least some alignment with each of the other species, but for the *rbcL-accD* spacer analysis only the *A. thaliana* sequence gave any significant alignment. For the purpose of comparison the average identity of the 31 *B. napus* plastid coding sequences is also included (taken from Figure 2.5).

<table>
<thead>
<tr>
<th></th>
<th><em>A. thaliana</em></th>
<th><em>N. tabacum</em></th>
<th><em>S. oleracea</em></th>
<th><em>O. elata</em></th>
<th><em>Z. mays</em></th>
<th><em>O. sativa</em></th>
<th><em>P. thunbergii</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average coding seq. identity</td>
<td>97.8%</td>
<td>89.6%</td>
<td>88.41%</td>
<td>86.17%</td>
<td>83.42%</td>
<td>83.11%</td>
<td>73.25%</td>
</tr>
<tr>
<td><em>trnV-rps12</em> spacer</td>
<td>93.7%</td>
<td>68.7%</td>
<td>67.0%</td>
<td>70.4%</td>
<td>58.5%</td>
<td>58.4%</td>
<td>23.3%</td>
</tr>
<tr>
<td><em>rbcL-accD</em> spacer</td>
<td>80.4%</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 2.9. Intergenic Region Sequence Similarities. Percentage nucleotide identity of the *B. napus* plastid intergenic sequences with their homologues in the seven other species obtained from GenBank. The *B. napus* *trnV-rps12* 3’ spacer yielded at least some alignment with each of the other species, but for the *rbcL-accD* spacer analysis only the *A. thaliana* sequence gave any significant alignment. For the purpose of comparison the average identity of the 31 *B. napus* plastid coding sequences is also included (taken from Figure 2.5).

![Nucleotide Sequence Alignment of the Intergenic Spacer Region between the rbcL and accD Plastid Genes.](image)

**Figure 2.16.** Nucleotide Sequence Alignment of the Intergenic Spacer Region between the *rbcL* and *accD* Plastid Genes. 277 bp of the *B. napus* intergenic region was selected and aligned with the corresponding region from *A. thaliana*. The selection includes the *Acc1* restriction site, subsequently used as the heterologous gene-cloning site in the transformation vector pZBl. Deviations from the *B. napus* sequence, including substitutions, insertions and deletions are highlighted in yellow. Numbers in bold represent the nucleotide numbering allocated to the complete plastid sequence of *A. thaliana* in the GenBank database. The approximate position of the *accD* promoter/transcription start site is shown in blue.
### B. napus vs A. thaliana

<table>
<thead>
<tr>
<th>98%</th>
<th>96%</th>
<th>98%</th>
<th>97%</th>
<th>97%</th>
<th>72%</th>
<th>50%</th>
<th>79%</th>
<th>81%</th>
<th>27%</th>
<th>96%</th>
<th>82%</th>
<th>96%</th>
<th>97%</th>
<th>94%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start codon</td>
<td>Stop codon</td>
<td>Start codon</td>
<td>Stop codon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **rbcL coding sequence** (1440 bp in *B. napus*)
- **Intergenic sequence** (673 bp in *B. napus*)
- **accD coding sequence** (1470 bp in *B. napus*)

### B. napus vs N. tabacum

<table>
<thead>
<tr>
<th>90%</th>
<th>92%</th>
<th>92%</th>
<th>90%</th>
<th>75%</th>
<th>34%</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>44%</th>
<th>58%</th>
<th>68%</th>
<th>81%</th>
<th>82%</th>
</tr>
</thead>
</table>

**Figure 2.17. Sectioned Percentage Base Identities for the rbcL, accD and Intergenic Sequence.**

*B. napus* sequence fractions, between 135 and 300 bp in size, were independently aligned with the plastid genomes of both *Arabidopsis* (Sato *et al.*, 1999) and tobacco (Shinozaki *et al.*, 1986) to reveal various degrees of conservation. This spectrum is indicated on the illustration by both percentage base identities and a colour intensity scheme.
<table>
<thead>
<tr>
<th>Plastid Nucleotide Region</th>
<th>Sequence Identity between <em>B. napus</em> and <em>A. thaliana</em></th>
<th><em>K</em> value (Number of Substitutions per Site)</th>
<th>Nucleotide Substitution Rate = <em>K</em>/2<em>T</em></th>
</tr>
</thead>
</table>
| Average coding seq. (31 genes). | 97.8% | 0.022 | $0.507 \times 10^{-9}$  
 or  
 $0.76 \times 10^{-9}$ |
| *trnV-rps12* spacer | 93.7% | 0.063 | $1.452 \times 10^{-9}$  
 or  
 $2.179 \times 10^{-9}$ |
| *rbcL-accD* spacer | 80.4% | 0.196 | $4.518 \times 10^{-9}$  
 or  
 $6.778 \times 10^{-9}$ |

Table 2.10. Nucleotide Substitution Rates for Plastid Genes and Intergenic Regions. Sequence identity figures for *B. napus* and *A. thaliana* alignments were taken from Table 2.9 and the estimated divergence date (*T*) was taken to be 14.46-21.69 Myr (Table 2.7).
2.4. Discussion.

2.4.1. *B. napus* Plastid Genome Restriction Fragment Clone Libraries.

The *B. napus* libraries were primarily constructed as a starting point for the development of species-specific plastid transformation vectors. They were also used to sequence approximately 43% of the *B. napus* plastid genome. The libraries were generated by cloning restriction fragments of plastid DNA as was originally described for tobacco (Sugiura *et al.*, 1986). At the time these *B. napus* libraries were being constructed (1997/1998) the plastid genome sequence of the closely related *A. thaliana* (Sato *et al.*, 1999) was not yet available. Because the *N. tabacum* plastid genome was considered too divergent to be a useful guide to restriction fragment sizes in the *B. napus* genome, the decision as to which restriction enzymes to use for library construction was based on a restriction digest analysis of purified *B. napus* cpDNA (Figure 2.4). Initially it was decided to set up clone banks for two restriction enzymes *Hind*III (average fragment size approximately 5kb) and *Sal*I (average fragment size approximately 12kb), in the hope that any gaps in the former clone bank might be closed by overlapping *Sal*I fragment clones. Unfortunately 75% of the *Sal*I fragments and 26% of the *Hind*III fragments proved to be too large to clone into the pUC-based vectors used for library construction. For this reason, both *EcoR*I and *Xba*I fragments were also cloned, but these libraries were not pursued beyond the identification of clones covering any remaining gaps not filled by the *Hind*III and *Sal*I clones.

The assembly of cloned *B. napus* fragments into a plastid genome map was carried out by aligning their end sequences with the plastid genome sequence of tobacco (Shinozaki *et al.*, 1986; Ac. Z00044) and later confirmed by alignment with the plastid genome of *A. thaliana* (Sato *et al.*, 1999; Ac. AP000423). It was unfortunate that the *A. thaliana* sequence was not available sooner because it would have prompted the pursuit of the more practical *Xba*I library as opposed to the *Sal*I and *EcoR*I libraries which proved difficult and time consuming respectively. Despite the difficulties encountered the present clone bank has 97% plastid genome
coverage with significant overlap and it proved more than adequate for subsequent work.

2.4.2. Plastid Genome Sequence Analysis.

Sequencing the *B. napus* plastid genome was undertaken primarily with a view to assisting transformation vector design. A 9.4kb region surrounding the 16S rRNA gene (in the IR region) was sequenced to facilitate development of the vector pPC1 and 11kb in the *rbcL* region was sequenced to facilitate construction of the vector pZBl (see Chapter 3). 36.1kb was sequenced in the LSC region in order to provide additional data for phylogenetic studies and also to provide sequence information that would facilitate the construction of transformation vectors based on other intergenic integration sites in the future.

2.4.2.1. Coding Sequence.

Analysis of the assembled *B. napus* sequence data identified 29 protein-coding genes, 2 ribosomal RNA genes and 15 transfer RNA genes (56.6kb in total). The gene order was exactly that found in the plastid genome of *A. thaliana* (Sato *et al.*, 1999), demonstrating the close evolutionary relationship between the two species. The high degree of sequence similarity of the individual genes at both the DNA level (Table 2.5) and the amino-acid level (Table 2.6) further highlight this point. However, the most conclusive evidence comes from the neighbour-joining trees (Saitou & Nei, 1987; Figures 2.13 and 2.14) constructed from the alignment of concatenated amino-acid sequences. From these trees it is clear that *B. napus* and *A. thaliana* diverged more recently from each other than from the other selected higher plants including tobacco, spinach, evening primrose and the monocots.

2.4.2.2. Phylogenetic Analysis.

The dimensions of the angiosperm tree were used to estimate divergence dates for the various nodes on the tree (Tables 2.7 and 2.8). The absolute calibration point was taken to be 200 (± 40) million years since the monocot/dicot split (Wolfe *et al.*, 1989). This study concludes that *B. napus* and *A. thaliana* diverged from a common ancestor approximately 14.46-21.69 million years ago.
This estimation is close to the figure obtained by Yang et al. (1999a), who calculated the *Brassica-Arabidopsis* split at 14.5-20.4 Myr using a mitochondrial intron sequence taken from 10 species of Brassicaceae. With the knowledge that *Brassica* species are so closely related to the model plant *Arabidopsis*, it is clear that *Brassica* biotechnology will benefit from availability of the entire *Arabidopsis* nuclear genome (The *Arabidopsis* Genome Initiative, 2000).

The divergence of rosids from asterids was calculated to be 85.71-128.57 Myr in this study. This is a slightly younger age than that predicted by a mitochondrial intron comparison, placing the divergence at 112-156 Myr (Yang et al., 1999a). The calibrated “plastid multi-gene” tree also yielded dates of 84.90-127.35 Myr for the spinach-tobacco split and 80.93-121.40 Myr for the primrose-Brassicaceae divergence. The fact that the latter three dates are in such a close proximity to each other is indicative of the explosive angiosperm species radiation, believed to have occurred approximately 100 million years ago (Lidgard and Crane, 1988). This study has also positioned the maize-rice split at 29.39-44.08 Myr. Both these monocot species belong to the Poaceae family, a clade understood to have arisen more than 70 million years ago (Bremer, 2000).

Although the plastid genome sequence of pine (Wakasugi et al., 1994) was used in the single-gene sequence identity analyses (Tables 2.5, 2.6 and 2.9), it was purposely excluded from the concatenated polypeptide multi-sequence alignments, which included only the angiosperm plastid sequence data. This decision was taken in order to maximise confidence in the divergence estimates for the more recent angiosperm lineages, such as those of *Brassica* and *Arabidopsis* (Figures 2.13 and 2.14).

Assigning a universal time-scale on the angiosperm tree relies on the assumption that the sequences evolved, roughly, at a constant rate. Unsurprisingly there is strong evidence pointing to the fact that there is considerable rate heterogeneity between (and within; see Figure 2.17) individual genes in the same genome, as well as between the different organellar genomes and species (Britton, 1986; Wolfe et al., 1987; Li, 1997). Chaw et al. (1997) for example observed that
the nucleotide substitution rate for the nuclear 18S rRNA sequence was considerably higher in maize than it was in rice. Because this analysis was confined to a single gene sequence, its potential for accurate phylogenetic conclusions was limited. In contradiction, the results of our analyses suggest that the average nonsynonymous (amino acid changing) substitution rate for 27 plastid genes is virtually identical for both rice and maize. Therefore conclusions drawn from single-gene analyses can be misleading. One of the more interesting features of the concatenated alignment approach was the way in which the tobacco plastid coding sequences appear to have a nonsynonymous substitution rate approximately twice that of spinach. The alignment itself was checked and found to contain no abnormalities, such as large gaps, that could cause gross malformation of the tree. Future analysis of sequence divergence rates for these two and other related asterid species will hopefully explain this result.

All of the above conclusions are supported by the fact that they are derived from the phylogenetic information contained in 26 genes, encompassing various product functions, as opposed to a single gene. Support measures such as bootstrap values (Felsenstein, 1985) and parsimony jackknifing (Farris et al., 1996) have been shown in previous studies to be greatly improved using multiple genes (Soltis et al., 1999). However, such values were also shown to be reduced in studies involving a relatively small number of taxa and when all of the genes were taken from only one of the three alternative genomic compartments (Barkman et al., 2000). The bootstrap values (Felsenstein, 1985) calculated for the angiosperm clades (Figure 2.13) provided considerable support for the main branches of the tree.

With the knowledge that the *B. napus* nuclear genome arose from a hybridisation involving relatively recent ancestors of *B. oleracea* (made up of numerous subspecies including cauliflower, cabbage and broccoli) and *B. rapa* (turnup), it is clear that its chloroplast genome must also have been sourced from one of the two species (Paterson et al., 2001). Unfortunately, the current lack of chloroplast genome sequence for either *B. oleracea* or *B. rapa* makes it difficult to determine the exact origins of the *B. napus* chloroplast genome. However, once
these sequences do become available this question will be answered. This will have notable significance to the Brassica napus-specific transformation vectors pPC1 and pZB1 (see Chapter 3), as it will extend their application to other related species.

2.4.2.3. Nucleotide Substitution Rates.

The nucleotide substitution rate for the average plastid coding sequence was calculated using Kimura's two-parameter method (Kimura, 1980; Table 2.10). A rate of $0.507-0.76 \times 10^{-9}$ substitutions per site per year was obtained from the acquired knowledge that B. napus diverged from A. thaliana 14.46-21.69 Myr ago (Table 2.7) and a nucleotide sequence identity of 97.8% for the average plastid gene (Table 2.5). Kimura's method was chosen because it can allow a direct comparison of substitution rates between protein-coding and non-coding regions. The reason why this tends not to be the method of choice when comparing substitution rates between protein coding genes is that it does not distinguish between synonymous and nonsynonymous substitutions, leaving the possibility for a biased estimation. Methods such as that of Li et al. (1985) can correct for this bias but do not allow for comparisons with non-coding sequence.

The nucleotide substitution rate for the trnV-rps12 3' spacer sequence is $1.452-2.179 \times 10^{-9}$ substitutions per site per year (Table 2.10), almost 3 times greater than the genic rate. The rate for the rbcl-accD spacer is $4.518-6.778 \times 10^{-9}$ substitutions per site per year, 9 times faster than the genic substitutions. It is clear that the trnV-rps12 3' intergenic region is mutating considerably slower than the rbcl-accD region. It is possible that this is a consequence of its location in the inverted repeat region of the plastid genome. A previous report has shown that IR sequences have lower rates of nucleotide substitution than the rest of the plastid genome (Wolfe et al., 1987). The sequence identity values for the five IR genes contained in this study (rrn16S, rrn23S, ndhB, rps7 and rps12 3'; Tables 2.5 and 2.6) further support this claim.
2.4.2.4. Intergenic Sequence Analysis.

The primary reason for the analysis of the plastid intergenic spacer sequences was to facilitate the design and construction of plastid transformation vectors and to determine the extent to which such sequences would confer species-specificity on their use. The Brassica-specific transformation vector pPCI (see Chapter 3) was constructed to facilitate the insertion of heterologous DNA into the trnV-rps12 3' intergenic region via homologous recombination. The XbaI site found mid-distance between the two coding regions was chosen as the polylinker cloning site on the basis that it is located in a region that can tolerate a high level of mutation (Figure 2.15). Therefore it is unlikely that insertion of a foreign gene at this point will interrupt the function of neighbouring plastid genes. The XbaI site was chosen in preference to the AflII site (412 bp closer to the rps12 3' coding sequence), as it is unique to the restriction fragment clone Hin18 (unlike AflII) making vector construction more straightforward. This AflII site is conserved in Solanum nigrum and was used in the Solanum-specific chloroplast transformation vector pSSH1 as the location for a multiple cloning site (polylinker) for insertion of transgenes (Kavanagh et al., 1999).

During the construction of the transformation vector pZBl (see Chapter 3) the heterologous gene insertion site chosen was AccI, found 310 bp upstream of the accD start codon. This is approximately 190 bp closer to accD than the cloning site used in the N. tabacum plastid transformation vector pZS197 (Svab and Maliga, 1993). Although closer to accD, concern about possible promoter interference was eased by the knowledge that, for tobacco at least, the accD gene is transcribed exclusively from a NEP promoter 129 bp upstream of the start codon. NEP consensus promoters are typically confined to a small (approximately 10 bp) stretch of DNA sequence adjacent to the transcription start site (Hajdukiewicz et al., 1997). Assuming transcription in B. napus starts at approximately –129 bp like in tobacco, the additional 180 bp distance further upstream to AccI should be sufficient to avoid any problems. Numerous insertions and deletions in the rbcL-accD intervening sequence, observed in the alignment of B. napus sequences with those of A. thaliana (Figure 2.16) add further weight to the conclusion that this region of the plastid genome in devoid of any transcriptional function. If it did
have any such function then a level of nucleotide sequence conservation as high as that observed in the coding regions (Table 2.7) might have been expected. Interestingly sectioned identities covering this region (Figure 2.17) imply that even the gene-coding sequences immediately flanking the spacer region (both for 3' of \textit{rbcL} and 5' of \textit{accD}) are less conserved than the remainder of the gene. This appears to be particularly true for \textit{accD}, a gene noted for the fact that it has been eliminated from the plastid genome of monocot species (Maier \textit{et al.}, 1995).
3.1. Introduction.

3.1.1. Plastid Transformation Vectors.

A typical plastid transformation vector consists of (a) a standard *E. coli* plasmid replicon, (b) chloroplast DNA targeting sequences that act to direct transgene integration by homologous recombination, and (c) selectable markers that function in *E. coli* and in chloroplasts. The utilisation of a high copy number *E. coli* plasmid such as those of the pUC variety (Yanisch-Perron *et al.*, 1985) facilitates the preparation of sufficient quantities of vector DNA. Because foreign DNA integrates into the plastid genome exclusively via homologous recombination (Kavanagh *et al.*, 1999), the point of integration is determined by the choice of flanking plastid DNA sequence. Guidelines governing the length of this sequence are not very well defined. Efficient recombination has been achieved using between 3kb (Svab and Maliga, 1993) and 7.8kb (Kavanagh *et al.*, 1999) of homologous targeting sequence. The choice of selectable marker strategy is also to some extent dependant on the region of plastid genome targeted by the transformation vector. If for example, selection is based on mutant alleles of chloroplast rRNA and ribosomal protein genes (which confer "binding-type" antibiotic resistance), then these sequences can be used not only to select for transplastomic shoots but also to direct foreign gene integration (see General Introduction 1.4.3.1). When the flanking plastid sequences in a transformation vector cannot themselves confer resistance, a heterologous marker gene such as the bacterial *aadA* gene, which confers spectinomycin and streptomycin resistance, may be included in the vector.

Transcription of foreign transgenes is typically controlled by the inclusion of a native plastome promoter upstream of the gene. Both the constitutive rRNA operon promoter (*Prrn*; Svab and Maliga, 1993) and the *psbA* promoter (*PpsbA*;
Zoubenko et al., 1994) have been successfully used to drive foreign gene expression in transplastomic tobacco. Translational efficiencies can also be improved by the use of plastome-derived ribosome-binding sites and 3'—regulatory elements. The 3'—terminator sequences of the rps16 and psbA genes have been used to stabilise foreign gene transcripts (Zoubenko et al., 1994).

Probably the most critical decision to be made when constructing a plastid transformation vector concerns the choice of plastid targeting sequences and the transgene insertion point. Ideally, the location of the transgene should not interfere in any way with the expression of neighbouring plastid genes (unless the aim of transformation is gene knock-out; see General Introduction 1.4.4). For this reason intergenic regions are considered the “safest” locations for transgene expression cassettes. Several non-coding intergenic regions have been targeted in tobacco and other higher plants and the resulting transformants exhibited no morphological abnormalities that might signify altered plastid gene expression. Two vectors, named pZS197 (spectinomycin selection; Svab and Maliga, 1993) and pTNH32 (kanamycin selection; Carrer et al., 1993), were developed for the site-specific integration of foreign sequences between the rbcl and accD (ORF512) genes located in the Large Single Copy region of the tobacco plastid genome. The pPRV vector series targets insertion between the divergently transcribed trnV and rps12 genes in the inverted repeat (IR) region of the tobacco plastome (Zoubenko et al., 1994). The pSSH1 vector of Kavanagh et al. (1999) targets the same region using plastid sequences of Solanum nigrum L. Koop et al. (1996) chose the non-coding sequence between rpl32 and trnL in the Small Single Copy region as a transgene insertion site. Foreign genes have also been inserted in the spacer sequence between the trnl and trnA genes in the IR (Daniell et al., 2001) and also between the trnfM and trnG genes in the Large Single Copy region, adjacent to the rpo operon (Ruf et al., 2001; Note: Chapter 2 describes the sequence of a homologous region in B. napus). Interestingly the latter site was selected on the basis that it resulted in particularly high chloroplast transformation frequencies.
3.1.2. Plastid Transformation Efficiencies.

Stable plastid transformants are now routinely obtained in tobacco at a frequency of one transformant per bombarded leaf using the heterologous \textit{aadA} gene as the selectable marker (Svab and Maliga, 1993). This frequency is reported to be approximately 100 times greater than that noted for transformants recovered using the spectinomycin resistance mutations in the 16S rRNA gene (Svab \textit{et al.}, 1990). However, the generalisation that the use of the dominant \textit{aadA} is the superior method for spectinomycin selection is not supported by results obtained using the polyethylene glycol-mediated protoplast transformation method. Using the \textit{aadA} marker, Koop \textit{et al.} (1996) obtained 2-4 plastid transformants for every $10^5$ treated protoplasts. Kavanagh \textit{et al.} (1999) received a similar frequency of one transformant out of $10^5$ PEG-treated protoplasts, using mutant ribosomal genes as the basis of selection. Therefore, the use of recessive binding-type selectable markers should not be dismissed, especially since the resulting transformants typically reach the homoplasmic state after a single round of regeneration under selective conditions (Kavanagh \textit{et al.}, 1999).

A chimeric \textit{kan (npt II)} gene, encoding neomycin phosphotransferase, has also been used to select for plastome transformants following biolistic bombardment (Carrer \textit{et al.}, 1993). Unfortunately, its efficiency makes it less attractive than the non-lethal spectinomycin resistance marker. Only one transplastomic clone was obtained per 25 bombarded leaf samples. Also a significant background of nuclear transformants complicated screening steps. The quest for superior plastid transformation efficiencies has led to the recent development of a selection system based on chimeric betaine aldehyde dehydrogenase (BADH) expression facilitating the detoxification of betaine aldehyde in the regeneration medium. Daniell \textit{et al.} (2001) demonstrated as many as 23 transformants could be obtained from a single bombarded leaf disk using the BADH gene (see General Introduction 1.4.3.2).

It is still unclear whether or not the size of the targeting fragments (for homologous recombination) influences plastid transformation efficiency. Zoubenko \textit{et al.} (1994) reported a five-fold increase in efficiency when the
sequences flanking the $aadA$ cassette comprised $1.5 + 3.6$kb as compared with $1.8 + 1.2$kb. However, Koop et al. (1996) found similar transformation efficiencies for two vectors whose targeting sequences differed not only in size but also in their plastome insertion point.

The degree of divergence between the donor (vector) and target plastid sequences may also influence the transformation efficiency. Newman et al. (1990) described a 10-100-fold decrease in integration frequencies when transforming the *Chlamydomonas reinhardtii* plastome with a partially homologous *C. smithii* targeting sequence. Interestingly, the situation appears to be less stringent in higher plant plastid transformation. Kavanagh et al. (1999) successfully transformed the tobacco plastome with plastid DNA from *S. nigrum*. The 7.8kb *Solanum* sequence was on average 2.4% divergent from the corresponding tobacco plastome region, yet the frequency of transformants obtained was similar to that observed for other PEG-mediated transformation experiments that utilised completely homologous donor DNA (Golds et al., 1993; Koop et al., 1996). This data suggests that the RecA-mediated mismatch repair system may be naturally suppressed in higher plant chloroplasts. It also suggests that it may be possible to utilise a “universal vector” for chloroplast transformation in a number of different higher plant species. Sequence analysis of the tobacco plants transformed with pSSH1 revealed multiple recombination events (Kavanagh et al., 1999), which differs from the mechanism of plastid transformation in *Chlamydomonas*, where only a single double recombination event leads to integration of the donor DNA (Newman et al., 1990).

### 3.1.3. Plastid Transformation in Higher Plants other than Tobacco.

Unfortunately, twelve years on from the first reported case of higher-plant chloroplast transformation (Svab et al., 1990), tobacco remains the only species to be successfully transformed on a routine basis. Recent years have seen much effort being placed into expanding the technology to other more commercially useful crops. Stable plastome transformants were generated for both *Arabidopsis* (Sikdar et al., 1998) and potato (Sidorov et al., 1999), but disappointingly only infertile transformants were regenerated. Sidorov et al. (1999) used an optimized tobacco regeneration and selection protocol in order to obtain potato transformants. Gentle
wounding of the leaf surface was found to benefit the regeneration response. Despite such efforts, a frequency of about one potato transformant per 15-30 bombardments was obtained, an order of magnitude less than that reported for tobacco (Svab and Maliga, 1993). However, the fact that the transformation vectors used to bombard the potato tissue were based on tobacco plastome sequences and not those of potato (both species belong to the Solanaceae; Angiosperm Phylogeny Group, 1998) suggests the possibility that this frequency might be improved by using a potato-specific vector (DeGray et al., 2001).

Most cereal crops are naturally resistant to spectinomycin (Fromm et al., 1987). Transformation of non-green rice plastids was achieved using a chimeric marker gene coding for a fusion protein consisting of a GFP domain (which acted as a visual selectable marker) and an aminoglycoside 3'-adenyl-transferase domain (which conferred spectinomycin resistance) (Khan and Maliga, 1999). Embroygenic tissue was chosen for bombardment because it has a higher regeneration potential than leaf-tissue in graminaceous species (Brownlee and Berger, 1995; Bilang and Potrykus, 1998). Although a reasonable transformation frequency was obtained (12 out of 25 bombardments), all transformants were found to be heteroplastic and contained sectors of both wild-type and transformed tissue that could be identified on the basis of GFP fluorescence.

The recent transformation of the tomato chloroplast genome can be looked upon as a huge step forward in the application of transplastomic technology because it is the first agronomically important crop species to yield fertile homoplastic transformed plants (Ruf et al., 2001). In spite of its relatively recent ancestry with tobacco (both species belong to the Solanaceae, which also includes potato) the tomato transformation protocol differs greatly. It was found that the use of lower light intensities (than those used for tobacco) was critical for the successful regeneration of transplastomic tomato callus cells. Also the technique described was relatively time-consuming because of the need to first select for homoplastic undifferentiated callus before inducing the regeneration of shoots. Despite the need to optimise the technique and the use of a tobacco-specific
transformation vector, it nevertheless provided a workable yield of approximately one stable chloroplast transformant per 10 bombardments.

From these few examples it is clear that if the plastid genome of *B. napus* is to be successfully transformed, the standard tobacco transformation technique (Svab *et al*., 1990) will almost certainly have to be modified. Moreover, because of the extent of sequence divergence between the plastomes of *N. tabacum* and *B. napus* (Chapter 2), it would be unwise to attempt transformation of the latter using plastid sequences from the former. Currently, the most efficient method of *B. napus* nuclear transformation involves *Agrobacterium* infection of hypocotyl tissue. Transformants have been successfully regenerated from hypocotyl explants (Radke *et al*., 1988) and efficient shoot regeneration can also be induced using protoplasts isolated from the hypocotyls of *B. napus* cultivars (Cheng *et al*., 1994; Skagen and Iversen, 1999).

This Chapter describes the construction of two types of vector suitable for transformation of the plastid genome of *B. napus*. The first was engineered to contain point mutations in the *rrn16* and *rps12-3’* genes which confer “binding type” resistance to both spectinomycin and streptomycin while the second contains the bacterial *aadA* gene which confers “antibiotic detoxifying” resistance.
3.2. Materials and Methods.

3.2.1. Materials.

3.2.1.1. Plasmids.

pUC7 (Messing and Vieira, 1982) was used for the construction of both the *B. napus* chloroplast transformation vectors pZB1 and pPC1.

pJDH118 (Hoheisel, 1989) was used for a sub-cloning step during the construction of pPC1.

pUC19 (Yanisch-Perron, *et al*., 1985) was used for a sub-cloning step during the construction of pPC1.

pCP-aadA1 (unpublished; constructed by Alan Magee) was the source of the aminoglycoside 3′-adenyl-transferase (*aadA*) expression cassette used in the construction of pZB1.

3.2.1.2. *In vitro* Plant Material.

*Nicotiana tabacum* (tobacco) cv. Samsun NN was maintained under sterile conditions in MS medium (Murashige and Skoog, 1962) in a growth room under a 16 hour light/8 hour dark cycle at 23-25°C.

3.2.1.3. Oligonucleotides used for DNA Sequencing.

In addition to primers used during the *B. napus* plastid genome-sequencing project (section 2.2.1.5), a number of sequencing primers were designed especially for confirmation of various cloning and mutation introduction steps during the construction of the *B. napus* chloroplast transformation vectors pPC1 and pZB1.

To confirm the introduction of the spectinomycin point mutation, (Kavanagh *et al*., 1994; Fromm *et al*., 1987) during the construction of pPC1, the following primers were designed based on sequence already obtained for the 16S rRNA region:

138756Fc: 5′-CCT CAC CTT CCT CCG CGT TAT CAC C-3′
138857R: 5′-CGC GGA CAC AGG TGG TGC ATG GCT G-3′
The following primer was used to confirm a correct cloning junction during the construction of pPC1:

endNcoR: 5’-GAT CGA TCT CTT TCT CGA AAC-3’

The following primers were used to confirm correct cloning junctions during the construction of pZB1:

P19r: 5’-ATG CTT CCG GCT CGT ATG TTG TG-3’

EndEcoR: 5’-GCC AAG ATA TCA GTA TCC TTG G-3’

zbaccR: 5’-GAA ATA TTT GTT CTT CGC TTA TGA ATA-3’

3.2.1.4. Oligonucleotides used for PCR.

In addition to using the primer 101150F, used for sequencing of the 9.4kb *B. napus* region (see Section 2.2.1.5), the following primers were designed for PCR diagnosis of *Brassica* sequence insertion into the tobacco plastome via homologous recombination. Because of high sequence similarity, these primers were suitable both for tobacco and *Brassica* plastid sequence amplification:

SpecF: 5’-TCT CAT GGA GAG TTC GAT CCT GGC-3’

SpecR: 5’-AAA GGA GGT GAT CCA GCC GCA C-3’

102010R: 5’-TAG CAG GCA AAG ATT TCA CTC CGC-3’

PCR amplifications were carried out essentially as described as Sambrook *et al.* (1989) with 2.5mM MgCl₂ used in the reaction to increase yield. The thermal cycling parameters were 94°C for 1 min., 53°C for 1 min., 72°C for 1 min. for 30 cycles with a 5 min. extended polymerisation step at the end of the reaction.

The primer oligonucleotides used to introduce the spectinomycin C→T point mutation (Kavanagh *et al.*, 1994; Fromm *et al.*, 1987) by overlap extension PCR (Ho *et al.*, 1989) were:

16Sf: 5’-CTT GAT GTC ATC CTC ACC TTC CTC CGG C-3’

16Sr: 5’-TCA TCA TGC CCC TTA TGC CCT GGG CGA-3’
3.2.1.5. Oligonucleotides used to make Synthetic Polylinkers and Insertions.

Annealing the following oligonucleotide pair formed the synthetic polylinker, which was inserted into the AccI site in the intergenic region adjacent to the rbcl for the construction of pZBl:

**AcpolF:** 5'-ATA CGC GAT CGC CTA GGT CGA CGG GCT TAA GTG ATC ATT AAT TAA GTT TAA AC-3'

**AcpolR:** 5'-ATG TTT AAA CTT AAT TAA TGA TCA CTT AAG CCC GTC GAC CTA GGC GAT CGC GT-3'

The polylinker contains the following restriction sites: AccI, SgfI, AvrII, SalI, BclI, PciI, and Pmel.

The following oligos were annealed and inserted into pUC7 to facilitate the construction of pPC1:

**BBHECO1:** 5'-AAT TCC ATG GGG GAA GCT TGG GGA GCT CG-3'

**BBHECO2:** 5'-AAT TCG AGC TCC CCA AGC TTC CCC CAT GG-3'

During the construction of pPC1, introduction of the T→G point mutation conferring streptomycin resistance to a putative transformant (Kavanagh et al., 1994) was achieved by replacing an EcoRV-MscI fragment with the annealed oligos:

**STR1:** 5'-ATC TCA CAC CGG GTA AAT CCT GAA CCC TTC CCC CTC TTA CTA AGA CTA CAG AAT GTT CTT GTA AAT TAT GG-3'

**STR2:** 5'-CCA TAA TTT ACA AGA ACA TTC TGT AGT CTT AGT AAG AGG GGG AAG GGT TCA GGA TTT ACC CGG TGT GAG AT-3'

The annealing and insertion of the following pair of oligos into the intergenic XbaI site created the polylinker for pPC1:

**XpolA:** 5'-CTA GAG CGA TCG CGT CGA CGG TAC CGT TTA AAC-3'

**XpolB:** 5'-CTA GGT TTA AAC GGT ACC GTC GAC GCG ATC GCT-3'

The polylinker contains the following restriction sites: PmeI, KpnI, SalI, SgfI and XbaI.
3.2.2. Methods.

3.2.2.1. Overlap Extension PCR.

This PCR based technique for site-directed mutagenesis, developed by Ho et al. (1989), was used for the introduction of a point mutation into the 'pHE-700' construct. Complementary primers (one of which contains the base-change mutation, i.e. the mutagenic primer) were used to amplify the circular plasmid bidirectionally giving a linear product. This can be re-circularised by a simple ligation reaction.

A 50μl total PCR reaction mixture was prepared, containing approximately 500pg/μl plasmid template, 2ng/μl of each primer 16Sf and 16Sr, 0.1Unit/μl Pfu polymerase, optimised concentrations of MgCl₂, deoxy nucleoside triphosphate’s (200μM dATP, 200μM dCTP, 200μM dGTP, 200μM dTTP), 20mM Tris pH8.75; 10mM KCl, 0.1% Triton X, 10mM (NH₄)₂SO₄, 100μg/ml BSA. The thermocycle conditions used were 94°C (45 sec.); 60°C (45 sec.); 72°C (7 min., 20 sec.), put through 25 cycles and performed in a PTC-100™ Programmable Thermal Controller. Following gel-purification the linear fragment was self-ligated in accordance to the recommendations of Boehringer Mannheim. 2-20ng of this self-ligated fragment was then transformed into E. coli DH₅α cells. Transformants were selected on LB medium containing 100μg/ml ampicillin. The correctly constructed mutant plasmid was identified firstly by digestion analysis and then by sequencing.

3.2.2.2. Annealing Oligonucleotides.

Complementary single-stranded oligonucleotides were generally annealed in a medium salt concentration buffer (50mM NaCl, 10mM Tris-HCl, 10mM MgCl₂), with equal concentration of each oligo strand (approx. 2ng/μl). Solution was then placed in a boiling water bath (0.5L) and allowed to cool down to room temperature naturally.

3.2.2.3. ³²P Labelling of Oligonucleotides.

Oligonucleotide probes were used to screen colonies for their successful cloned insertion. Manufactured oligonucleotides are synthesized without a phosphate group at their 5' termini and are easily labelled by transfer of the γ-³²P
from [$\gamma^{-32}$P]ATP using bacteriophage T4 polynucleotide kinase. The reaction was carried out in accordance to the recommendations of the manufacturer (New England Biolabs). 30pmol [$\gamma^{-32}$P]ATP was added to 18.5pmol oligonucleotide and 5 units of polynucleotide kinase in a reaction volume 20$\mu$l. The remaining reaction buffer requirements were supplied as a 10X buffer concentrate, provided by the kinase manufacturer. The reaction was allowed to occur at 37°C for 30 minutes followed by incubation at 95°C for 5 minutes to denature the enzyme.

### 3.2.2.4. Plasmid DNA Maxiprep.

Plasmid DNA preps of high concentration and purity were derived using the kit obtained from QIAGEN. The cells from a 500ml overnight bacterial culture were recovered by centrifugation for 15 minutes at 6000 rpm in a Sorval RC-5B centrifuge (GSA rotor) at 4°C. In accordance to the manufacturers’ protocol the pellet was suspended in the buffers provided. To improve plasmid yield the DNA solution was eluted from the Qiagen-tip® into a number of fractions. Each fraction was checked for DNA content by electrophoresing a sample on 0.75% agarose stained with Ethidium Bromide. Only those fractions containing higher concentrations of plasmid were retained and pooled prior to ethanol precipitation. Air-dried pellets were suspended in 200$\mu$l sterile H$_2$O and typically produced plasmid concentrations in excess of 1 $\mu$g/$\mu$l.

### 3.2.2.5. Preparation of Tungsten Particles for Biolistic Transformation.

1ml of ice-cold 70% special grade ethanol was added to 30mg of accurately weighed tungsten powder (M-17, average particle size 1.1$\mu$m; Alpha) and the suspension was vortexed vigorously for 5 min. followed by 15 min. soaking on ice. The tungsten was then gently sedimented by spinning at 8000 rpm for 5 seconds in a microfuge and the ethanol was removed using a Gilson. The tungsten particles were then washed three times by adding 1ml of sterile doubly distilled H$_2$O, vortexing vigorously for 1 min and microfuging at 8000 rpm for 3 seconds before removing the H$_2$O. During the third wash a further 3 second 8000 rpm spin was performed to ensure as much as possible of the H$_2$O was removed. 500$\mu$l of sterile
50% glycerol was added and the suspension was stored at -20°C until use for a maximum period of one week.

3.2.2.6. Coating Washed Tungsten Particles with DNA.

The glycerol stock was vortexed vigorously for five minutes and 50μl was transferred to an autoclaved 1.5ml eppendorf tube before the tungsten had a chance to settle. While vortexing vigorously the following were added to the tungsten aliquot in the order given: (1) 5μl of a 1-2μg/ml plasmid DNA solution; (2) 50μl of 2.5M CaCl2 solution (3) 20μl of 0.1M spermidine (free base; tissue culture grade) solution. The plasmid DNA maxiprep was microfuged for 5 min. at 15000 rpm immediately before use to sediment residual insoluble matter. The CaCl2 solution was prepared fresh, filter sterilised and stored on ice before use. The spermidine solution was prepared from a 1M stock solution that had been filter sterilised and stored at -20°C for a period no longer than one month. The tungsten particles were then vortexed vigorously for 30 min. at 4°C. After vortexing the particles were allowed to settle for 1 min. before being sedimented by microfuging for 2 seconds at 8000 rpm. The liquid was removed and discarded and 140μl of 100% special grade ethanol was added. The ethanol was removed without disturbing the sedimented tungsten and 48μl of 100% ethanol was added. The tungsten was gently resuspended by tapping the side of the tube and by low speed vortexing. The DNA coated particles, stored on ice until use, were used within a few hours of preparation and were resuspended by gentle vortexing before loading on the flying disc (macrocarrier).

3.2.2.7. Biolistic Transformation of Tobacco Leaf Tissue.

Healthy tobacco leaves 3-5cm in length from plants 4-5 weeks old were removed under sterile conditions and the abaxial side was bombarded. Biolistic transformation was carried out as described in the BioRad manual for the DuPont PDS1000/He Particle Delivery System. The following settings were used for plant leaf transformation unless otherwise stated;

(1) Helium pressure at the tank regulator: 1300-1400 psi.

(2) 1100 psi rupture discs were used.
(3) Flying disc assembly: level 2 from the top.
(4) Flying disc (macrocarrier) was 6mm from the stopping screen.
(5) Petri dish holder: level 2 from the bottom.
(6) Vacuum at time of shot: 28 in. Hg.
(7) The leaf was placed in a Petri plate containing MS or RMOP medium to
    cushion the blast.

Following bombardment the petiole was pushed into the medium and the
plate was sealed and stored in the growth room for two days.

3.2.2.8. Regeneration of Stable Transformants.

Following the two day incubation in the growth room on RMOP medium in
the absence of selection the bombarded leaves were cut into 5mm² sections and the
bombarded side was placed on shoot regenerating RMOP (Svab et al., 1990)
medium containing 500mg/L spectinomycin. Resistant tissue that appeared after 6-
8 weeks was put through a further round of shoot regeneration before the shoots
were rooted in MS medium containing spectinomycin 500mg/L.

Spectinomycin resistant plantlets were tested for streptomycin resistance by
rooting in MS medium containing both spectinomycin 500mg/L and streptomycin
500mg/L.
3.3. Results.

3.3.1. Construction of *B. napus* Plastid Transformation Vectors.

Having generated clone banks covering 97% of the *B. napus* plastid genome (Chapter 2), the choice of flanking sequence to be used in the construction of transformation vectors was virtually unrestricted. Two vectors were constructed: one was designed to direct integration of foreign transgenes into the non-coding region located between the *trnV* and *rps12* (3’) genes in the Inverted Repeat (IR) region; the other to direct integration into the *rbcL-accD* intergenic spacer, located in the Large Single Copy region. Both insertion points were chosen on the basis that these integration sites had been successfully used for plastid transformation in other higher plant species (Zoubenko *et al*., 1994; Kavanagh *et al*., 1999; Svab and Maliga, 1993).


Because the *trnV-rps12* (3’) intergenic sequence is located beside ribosomal genes selection can be based on mutant alleles of these genes, which confer “binding-type” antibiotic resistance (Kavanagh *et al*., 1994). Such selection systems have been successfully utilized in tobacco (Svab *et al*., 1990) and *S. nigrum* (Kavanagh *et al*., 1999) transformation vectors. The Brassica-specific vector pPC1 was designed to contain two “binding-type” selectable markers (rather than selection based on the bacterial *aadA* gene) because binding-type selection has been reported to increase the frequency of homoplasmy without the requirement for repeated rounds of regeneration (Svab *et al*., 1990). This strategy may be crucially important in the case of species which show low regeneration efficiencies *in vitro*.

The chloroplasts of most dicotyledonous species are naturally sensitive to the antibiotics spectinomycin and streptomycin. However, a number of specific point mutations within the 16S rRNA, 23S rRNA and *rps12* genes have been shown to confer resistance (Fromm *et al*., 1987; Kavanagh *et al*., 1994). Two such mutations were applied to the pPC1 transformation vector: a G→A transition
(eliminating an *AatII* restriction site) at position 1192 near the 3' end of the 16S rRNA which confers spectinomycin resistance; and an A→C transversion at the first nucleotide position of codon 87 of the *rps12* (3') gene (converting a lysine into a glutamine) which confers streptomycin resistance.

Molecular assembly of pPC1 involved numerous cloning steps, culminating in a construct consisting of approximately 6.4kb of chloroplast genome sequence into which was engineered (i) a multiple cloning site (ii) a point mutation conferring spectinomycin resistance and (iii) a point mutation conferring streptomycin resistance (as discussed above). Diagrammatic flow-charts describing each of these construction steps are provided in Figures 3.1, 3.2, 3.3 and 3.4. A circular map of the final transformation vector construct is shown in Figure 3.5. The *B. napus* plastid DNA sequences used in this construct were obtained from two of the previously cloned *HindIII* restriction fragments (clones Hin 18 and Hin 34; Figure 2.11 and Table 2.1) that encompass the 16S rRNA – *rps12* (3') coding region.

### 3.3.1.1.1. Introduction of a Spectinomycin Resistance Point Mutation.

The G→A base change conferring spectinomycin resistance was introduced into the *Brassica* 16S rRNA coding sequence by overlap extension PCR (Figure 3.1). This technique was developed by Ho et al. (1989) in order to perform site-directed mutagenesis at any site in a plasmid independent of restriction sites. Firstly, the smallest possible *B. napus* plastid DNA fragment containing the targeted nucleotide was subcloned from pHin 34 (Chapter 2). This was done in order to avoid unnecessary PCR amplification of flanking sequence, a process that can result in the incorporation of random mutations (hence the use of the proof-reading DNA polymerase *Pfu*, Cline et al., 1996). Digestion of the 6.4kb pHin 34 clone with both *HindIII* and *EcoR1* yielded 3 DNA fragments: 3kb, 2.7kb and 726bp). The 726bp fragment which contained the wild-type target nucleotide (G), was gel-purified and inserted into the pUC19 cloning vector (Figure 3.1). Correct assembly of the resulting construct ‘pHE-700’ was confirmed by the restriction digest analysis described in Figure 3.6.
pHE-700 was then subjected to PCR-amplification using two outward-facing primers: the '16Sf' mutagenic primer (containing the G→A base change which abolishes an AatII restriction site) and the '16Sr' primer. Amplification using these outward-facing primers resulted in the production of a linear plasmid-length (3.3kb) DNA product incorporating the desired point mutation (Figure 3.1). This mutant DNA product was re-circularized by "self-ligation". Only the circular form of the product was capable of being transformed and cloned in _E. coli_ DH5α. Correctly mutated clones were identified by screening for loss of the AatII restriction site (Figure 3.8) followed by confirmatory sequencing across the plastid DNA insert (Figure 3.9). Plasmid clones that yielded a single 3.3kb DNA fragment following AatII digestion (rather than two fragments of 2.65kb and 0.65kb) were considered to be candidate spectinomycin resistant mutants on the basis that the G→A base change eliminates an AatII site in the plastid DNA insert (a second AatII site is located in the pUC19 vector portion of the plasmid; Yanisch-Perron _et al._, 1985). However, a total of 28 clones each of which lacked the diagnostic AatII site had to be sequenced before a base-correct mutant 'pHE-700-ΔAatII' (Clone 21) was identified (Figure 3.9). Most clones were incorrect due to the exonucleolytic activity of _Pfu_ polymerase (Kroutil _et al._, 1996), which possesses a high rate of "primer-chewing" capability, which in turn caused small deletions in the ligated product (Figure 3.9).

A 508 bp plastid DNA fragment containing the spectinomycin point mutation was then gel purified from 'pHE-700-ΔAatII' following digestion with both _MluI_ and _EcoR1_. This mutant fragment was inserted into pHin 34 in place of the wild type homologous sequence. The resulting plasmid was called 'pHin 34-ΔAatII'. Figure 3.7 details the gel purification steps used to perform the latter cloning step as well as confirming the correct construction of 'pHin 34-ΔAatII' by restriction digest analysis. A single 6.4kb DNA fragment was obtained when 'pHin 34-ΔAatII' clones were digested with _AatII_, whereas two DNA fragments (5.3kb +1.1kb) were obtained when the parental pHin 34 was digested with _AatII_. Sequencing with the primers '138756Fc' and '138857R' (Figure 3.9) verified the introduction of the spectinomycin resistance mutation into 'pHin 34-ΔAatII' without any additional nucleotide changes.
3.3.1.1.2. Introduction of a Multiple Cloning Site.

The unique $X_{ba1}$ restriction site located in the trnV- rps12 (3') intergenic region was selected as the location for a multiple cloning site (polylinker) into which foreign transgenes could be introduced, based on the conclusions derived from the sequence analysis discussed previously (see Section 2.4.2.4). A 33 bp synthetic polylinker was designed to introduce four restriction endonuclease sites ($P_{mel}$, $K_{pnl}$, $S_{alI}$ and $S_{gfl}$) adjacent to the intergenic $X_{ba1}$ site. All sites in the polylinker were selected on the basis that they would be unique in the final transformation vector and compatible with some of the commonly used plastid transgene expression cassettes. The *B. napus* restriction fragment clone pHin 18 was digested with $X_{ba1}$ and the polylinker (formed by annealing the complementary oligonucleotides ‘XpolA’ and ‘XpolB’) inserted by a ligation reaction (Figure 3.2). Transformed clones were screened initially by restriction digestion analysis (on the basis of an acquired $P_{mel}$ restriction site; data not shown) and then sequenced to determine the orientation of the polylinker and to confirm correct insertion of a single polylinker oligonucleotide (Figure 3.18). One such clone, termed ‘pHin 18-Xpol 5’, was selected for the following manipulations.

3.3.1.1.3. Introduction of a Streptomycin Resistance Point Mutation.

A T→G base change, conferring streptomycin resistance, was introduced into the *B. napus* rps12 (3') coding sequence by synthetic oligonucleotide replacement (Figure 3.3). The plasmid ‘pHin 18-Xpol 5' (which contains the polylinker described in the previous section) was digested with both $P_{stl}$ and $M_{scI}$ restriction endonucleases. This yielded two fragments, the smaller of which (765 bp) was subcloned into the cloning vector pJDH119. Confirmation of this cloning step is provided in Figure 3.10. It clearly demonstrates that the digestion of this subclone, termed ‘pJDH-ME’, with both $P_{stl}$ and $M_{scI}$ resulted in the excision of the 765 bp fragment from the vector. The *EcoRV* restriction site adjacent to the mutation target site is unique in the subclone, whereas the parent plasmid ‘pHin 18-Xpol 5’ contained several *EcoRV* sites. Digestion of ‘pJDH-ME’ using both *EcoRV* and *MscI* released a 71 bp fragment, which was replaced by a fragment formed by annealing the chemically synthesised oligonucleotides ‘STR1’ and
'STR2'. This molecule was completely homologous to the excised EcoRV-MscI fragment except for the T→G streptomycin resistance mutation. Sequencing with universal M13 primers (Figure 3.18) identified several successful transformants containing the desired mutation. Finally, following restriction digestion, the 765 bp PstI-MscI fragment containing the desired point mutation was gel-purified and recloned back into the parent ‘pHin 18-Xpol5’ replacing the homologous wild-type fragment. This plastid DNA clone, derived from pHin 18, now contained both the XbaI polylinker and the streptomycin resistance mutation. It was named ‘pHin 18-XPRPG’.

3.3.1.1.4. Completion of the pPC1 Construct.

The exact length of plastid sequence included in the final transformation vector construct (to facilitate homologous recombination) was dependant on the positioning of the antibiotic resistant markers and the availability of suitable restriction sites. It was decided that the plastid flanking sequence would terminate at an NcoI site 1541 bp distant from the streptomycin resistance marker on one side and a SacI site 1384 bp distant from the spectinomycin resistance marker on the other (Figure 3.4). This meant that a total of 6450 bp B. napus plastid sequence was contained in the final pPC1 construct (Figure 3.5). To facilitate the required restriction site manipulations a special polylinker (‘BBHECO1’ annealed to ‘BBHECO2’) was inserted into the cloning vector pUC7 to make a new cloning vector named ‘pUC-BBHECO’ (Figure 3.4). This step also ensured that no unnecessary sites would reside in the final pPC1 construct that might reduce its usefulness as a cloning vector.

pPC1 was assembled by cloning the previously modified plastid DNA fragments (i.e. modified to contain the resistance mutations and the polylinker) into this pUC-BBHECO via two sequential steps. Firstly (Step 1 in Figure 3.4), ‘pHin 34-ΔAatII’ was digested with both SacI and HindIII and the resulting 2088 bp fragment was gel purified and inserted into pUC-BBHECO. The resulting intermediate construct was termed ‘pUC-HS’. Secondly (Step 2 in Figure 3.4), ‘pHin 18-XPRPG’ was digested with both NcoI and HindIII and the resulting 4362 bp fragment was gel purified and inserted into ‘pUC-HS’ alongside the NcoI-
HindIII fragment. Ligation at the HindIII site restored the entire 16S rRNA gene sequence (with its spectinomycin point mutation). The agarose gels pictured in Figure 3.11 show the DNA fragments isolated by the gel purification steps described and also provide restriction digest confirmation for both the ‘pUC-HS’ and the pPC1 constructs. In both instances, the same sized fragment that was inserted was released following digestion with the cloning endonucleases.

In order to further confirm that the final pPC1 construct had been assembled correctly from all its component pieces, all of the cloning sites, the point-mutation markers and polylinker insertions were re-sequenced using the primers highlighted in Figure 3.4. Relevant portions of resulting sequence data are shown in Figure 3.18. The restriction analysis described in Figure 3.16 also revealed that all five of the foreign gene insertion sites (Pme1, Kpn1, Sal1, Sgf1 and Xba1) were unique to pPC1.

3.3.1.2. Construction of the Transformation Vector pZB1.

A second B. napus-specific plastid transformation vector was constructed, modelled on the tobacco plastid vector pZSl described by Svab and Maliga (1993) which incorporates a dominant antibiotic resistance gene marker. Because this vector contains the rbcL—accD intergenic region as a transgene insertion site, it does not have the option of utilising a “binding-type” selection system. Fortunately expression of the bacterial aadA gene (which confers both spectinomycin and streptomycin resistance) from this region of the plastome has already been shown to allow efficient plastid transformation in tobacco species (Svab and Maliga, 1993). Therefore it was the logical choice of selectable marker to be used in the B. napus-specific pZB1 transformation vector. Plastid-localised expression of the aadA gene was controlled by the inclusion of the ribosomal RNA operon promoter (Prrn) and the 3' untranslated region (UTR) of the plastid psbA gene (psbA 3'). Both expression elements have been studied in great detail (see General Introduction). Prrn drives high constitutive transcription and the psbA 3' terminator provides high transcript stability. Efficient translation was ensured by the inclusion of a synthetic ribosome-binding site in the 5' UTR of the aadA cassette. The functional specificity of such regulatory elements in chloroplasts...
reduces the likelihood of obtaining antibiotic resistant escapes caused by integration of the \textit{aadA} gene into the nuclear genome.

Previous analysis of the \textit{rbcL–accD} intergenic sequence (see Section 2.4.2.4) highlighted the suitability of the \textit{AccI} restriction site for transgene insertion. The entire intergenic sequence was contained in the \textit{HindIII} restriction fragment clone pHin 32 (Figure 2.11 and Table 2.1). Unfortunately only part (the 3' 494 bp) of the \textit{rbcL} gene sequence was contained in this particular clone and if used alone would possibly be of insufficient length to achieve efficient homologous recombination. The remaining \textit{rbcL} sequence was located at one end of the plastid DNA \textit{HindIII} fragment cloned in pHin 27 (see Chapter 2). Analysis of both clones revealed a number of restriction sites not present in the plastid DNA sequence that could be incorporated into the transformation vector for use as a multiple cloning site (polylinker). The sites used were also selected on the basis of compatibility with existing plastid gene expression cassettes.

A diagrammatic flow-chart describing the cloning steps performed in the construction of pZBI is contained in Figure 3.12. A synthetic polylinker, formed by annealing the oligonucleotides “AcpolF” and “AcpolR”, was specifically designed in order to accommodate the introduction of the \textit{aadA} expression cassette (Svab and Maliga, 1993) and to possess multiple cloning sites flanking each side of the cassette. This polylinker was inserted into the unique \textit{AccI} site located between \textit{rbcL} and \textit{accD} in pHin 32 by ligating the kinased 53 bp oligonucleotides to the phosphatased vector digest. Successful transformants were diagnosed by \textit{SalI} digestion and sequenced to ascertain the orientation of the polylinker (Figure 3.19). The \textit{Prrn-aadA-psbA} 3' cassette was excised following both \textit{SalI} and \textit{AflII} digestion of the construct ‘pCP7-aadA1’ (provided by Alan Magee), which was gel-purified as a 1.3kb fragment and cloned into the \textit{SalI} and \textit{AflII} sites in the “Acpol” poly linker (‘pHin 32-Acpol’ digested with both \textit{SalI} and \textit{AflII}). Figure 3.14 provides evidence for the correct construction of ‘pHin 32-aadA’. The digestion of ‘pHin 32-aadA’ with both \textit{SalI} and \textit{AflII} resulted in the excision of the 1.3kb \textit{aadA} cassette fragment.
The final step in the construction of pZBl involved the simultaneous ligation of 3 DNA fragments: a 980 bp EcoRI-HindIII fragment containing the 5’ \textit{rbcL} sequence that was gel-purified from a pHin 27 digest; a 4.7 kp EcoRI-HindIII fragment containing the 3’ \textit{rbcL} sequence, \textit{aadA} cassette and \textit{accD} gene that was gel-purified from a ‘pHin 32-aadA’ digest; and the pUC7 cloning vector, digested with EcoRI. An agarose gel containing EcoRI-HindIII digests of pHin 27, ‘pHin 32-aadA’ and pZBl is shown in Figure 3.15. It shows that pZBl contains the 4.7kb fragment derived from ‘pHin 32-aadA’ and the 980 bp fragment derived from pHin 27. To comprehensively ensure the accuracy of the pZBl construct, all of the cloning junctions used during its assembly were re-sequenced using the primers highlighted in Figure 3.12. The restriction digest analysis described in Figure 3.17 confirmed the utility of the multiple cloning sites, showing each site to be unique to the plasmid construct. Figure 3.13 provides a circular genetic map of the pZBl plastid transformation vector.

3.3.2. Transformation of the Tobacco Plastid Genome using the \textit{B. napus}-specific vector pPC1.

The lack of an operational system for achieving chloroplast transformation in \textit{Brassica napus} meant that the most logical method for testing the vectors \textit{in planta} was via tobacco transformation. Despite the sequence divergence existing between the vector and host sequences, the pPC1 plastid transformation vector was introduced into tobacco leaf chloroplasts by biolistic bombardment. A total of twelve tobacco leaves (var. Samsun) were bombarded with tungsten particles coated with pPC1 DNA and five leaves were bombarded with pUC19 to act as a DNA control. The leaves were incubated for two days using non-selective growth conditions. Of the pPC1-bombarded leaves, ten were aseptically cut into 5mm\(^2\) sections and plated in shoot regeneration medium containing 500mg/L spectinomycin. The remaining two were also cut but instead plated in shoot regeneration medium lacking spectinomycin. Of the pUC19-bombarded leaves, three were cut and placed on spectinomycin-containing regeneration medium, while two were plated on medium without antibiotic selection. After approximately eight weeks prolific uncontrolled growth was observed from both the pPC1 and pUC19 bombarded leaf disks regenerated on antibiotic free medium.
(Figure 3.20 A and B). No green tissue (shoots or callus) regenerated from any of the pUC19-bombarded leaf sections plated on spectinomycin-containing medium (Picture C). Four independent shoots (Picture D) and a single green callus growth were found growing from the bombarded surface of leaf section that was in contact with spectinomycin-containing medium.

The five putative transformants were propagated and placed on regenerating medium containing 500mg/L spectinomycin for a second round of regeneration. Following three weeks incubation in optimal growth room conditions three of the four shoots remained green and developed shoots on the spectinomycin. These three clones, named PCT1, PCT2 and PCT3, were cut and placed in MS medium containing 500mg/L spectinomycin. The other one of the shoots and the green callus turned white and did not develop roots, even after eight weeks. PCT1, PCT2 and PCT3 were also cut and placed in MS medium containing both 500mg/L spectinomycin and 500mg/L streptomycin. All three shoots bleached and did not develop roots after eight weeks, clearly indicating their streptomycin sensitivity.

3.3.2.1. Molecular Analysis of the Putative pPCl Transformants.

In order to determine whether the spectinomycin resistant tobacco shoots were genuine pPCl transformants PCR analysis was performed on purified chloroplast DNA. The primers ‘SpecF’ and ‘SpecR’ were designed to amplify the plastid 16S RNA gene sequence, irrespective of whether it was from tobacco or Brassica sequence. A PCR product of approximately 1.5kb was obtained for each of the PCT1, PCT2, PCT3, wild type tobacco and wild type B. napus templates (data not shown). Digesting each PCR fragment with AatII revealed that while both wild type fragments and that of the PCT2 clone were cut into two pieces, those of PCT1 and PCT3 remained uncut (Figure 3.21). The persistence of the AatII site in the PCT2 ruled it out as being a genuine plastid transformant, suggesting it might be an “escape” or a novel spontaneous spectinomycin resistant mutant. The region surrounding the spectinomycin resistance point mutation on the PCR fragments was sequenced using the ‘138857R’ primer to determine whether it was homologous to B. napus or tobacco. Figure 3.22 shows the PCT1 PCR
sequence aligned with the homologous sequence in *N. tabacum*. It clearly highlights the occurrence of two additional base changes located on either side of the spectinomycin resistance point mutation. These base changes were compatible with the plastid *rrn 16S* sequence in *B. napus* (see Chapter 2). The occurrence of an otherwise homologous tobacco sequence would have implied that the spectinomycin resistance mutation most probably arose by spontaneous mutation.

The PCT1 and PCT3 clones were then checked to see if they contained the *XbaI* polylinker in the *trnV-rps12* 3' intergenic region. PCR primers, homologous to both *Brassica* and tobacco plastid sequences, were used to amplify the DNA region that might contain the polylinker. The agarose gels pictured in Figure 3.23 contained the PCR products from both the PCT1 and PCT2 templates in addition to those from wild-type tobacco and *B. napus* and also the purified plasmid pPC1 (controls). The high degree of sequence divergence that exists between tobacco and *Brassica* in non-coding regions meant it was not surprising to see a notable difference in the size of their respective PCR products. It was observed that the PCR fragments amplified from both the PCT1 and PCT3 clones were of similar size to the tobacco wild type (820 bp). Restriction digest analysis of the PCR fragments with both *Pmel* and *XbaI* confirmed that neither clone had incorporated the multiple cloning sites from the transformation vector.
Figure 3.1. Cloning Flow-Chart for the Introduction of a Spectinomycin Resistance Point Mutation into the *Brassica napus* Chloroplast Fragment Clone Hin 34. The C→T base change was achieved via overlap extension PCR and diagnosed by the loss of an AatII site.
Figure 3.2. Cloning of the artificially synthesised Polylinker into the XbaI Restriction Site in the rps12 3' - trnV intergenic Region.
Figure 3.3. Cloning Flow-Chart for the Introduction of a Streptomycin Resistance Point Mutation into the *Brassica napus* Chloroplast Fragment Clone already containing the *XbaI* polylinker. T to G base change was achieved by replacing a small restriction fragment with a synthetic dinucleotide fragment formed by annealing the oligonucleotides STR1 and STR2.
Figure 3.4. Final Cloning Steps for the Construction of pPC1. The mutant Brassica clone fragments were spliced into the pUC based vector with specially designed polylinker. Junctions were verified by sequencing, the primers for which are represented on the diagram by red arrows. Note: not all plastid genes are labelled.
Figure 3.5. *Brassica napus* Chloroplast Transformation Vector pPC1. Constructed for the introduction of heterologous genes into the spacer region between *rps12 3‘* and *trnV* of the plastid inverted repeat region, via homologous recombination. The foreign gene multiple cloning site is flanked on either side by a different antibiotic resistance marker gene (Spectinomycin and Streptomycin) allowing for selection of the transformed plastid.
Figure 3.6. Restriction Digest Analysis for the Construction of the Subclone pHE-700. Lane 1: 1kb ladder; Lane 2: pHin 34 digested with both HindIII and EcoRI; Lane 3: pUC19 digested with both HindIII and EcoRI; Lane 4: Gel purified 726 bp fragment from the pHin 34 digest; Lane 5: pHE-700 digested with both HindIII and EcoRI.

Figure 3.7. Introduction of the Mutant MluI-EcoRI Fragment from the Subclone pHE-700-ΔAarII into the Parent pHin 34. Lane 1: 5.9kb DNA fragment was gel purified from pHin 34 digested with both MluI and EcoRI; Lane 2: 508 bp DNA fragment was gel purified from pHE-700-ΔAarII digested with both MluI and EcoRI; Lane 3: pHin 34 digested with AarII; Lane 4: 1kb ladder; Lanes 5-8: Putative pHin 34-ΔAarII digested with AarII.
Figure 3.8. Screening for Spectinomycin Resistance Marker Introduction by the Loss of Aat11 Restriction Site. **Lane 1:** Control; "pHE 700" plasmid digested with Aat11 restriction enzyme. Digest pattern consists of two bands, 2650bp and 650bp, resulting from the presence of two restriction sites (one in the plastid fragment sequence and the other in the cloning vector). **Lane 2:** "1kb ladder". **Lanes 3-10:** Putative "pHE 700 ΔAat11" clones digested with Aat11. The 3.3kb band is indicative of Aat11 cutting the plasmid only once, confirming the removal of one of the sites from the parent plasmid.

Figure 3.9. Sequence Analysis of the Spectinomycin Resistance Point Mutation, introduced by Overlap Extension PCR. Putative pHE 700 ΔAat11 clones were sequenced across the target nucleotide mutation marker from both directions using the primers 138756Fc and 138857R. Although the C→T base change was successfully introduced into most clones, sequence analysis revealed the occurrence of multiple nucleotide deletions. Such deletions were a result of the exonuclease activity of Pfu polymerase causing "primer chewing" to occur at the 5' ends of the primers used in the PCR (including clones 6 and 8). Fortunately clone 21 contained no deletion and was completely homologous to the wild type B. napus sequence, except for the C→T point mutation.
Figure 3.10. Restriction Digest Analyses of the Subclone pJDH-ME. Lane 1: pJDH119 digested with both PstI and MscI; Lane 2: pHin 18-Xpol 5 digested with both PstI and MscI; Lane 3: pJDH-ME digested with both PstI and MscI.

Figure 3.11. Restriction Digest Analysis for the Assembly of the Construct pPC1. Lane 1: pUC-BBHECO digested with both SacI and HindIII; Lane 2: 2.1kb DNA fragment, which was gel purified from pHin 34-AatII digested with both SacI and HindIII (Picture A); Lane 3: pUC-HS digested with both SacI and HindIII; Lane 4: 1kb ladder, Lane 5: pUC-HS digested with both Ncol and HindIII; Lane 6: 4362 bp DNA fragment, which was gel purified from pHin 18-XPRPG digested with both Ncol and HindIII (Picture B); Lane 7: pPC1 digested with both Ncol and HindIII.
Annealed oligos: AcpolF + AcpolR. Cloned into the unique AccI site of pHin 32.

aacD plastid expression cassette cloned into Acpol polylinker.

**Figure 3.12. Flow-Chart for the Construction of pZB1.** The cloning strategy was designed using the plastid sequence data obtained in Chapter 2. All cloning junctions were confirmed by sequencing, the primers for which are represented on the diagram by red arrows.
Figure 3.13. *Brassica napus* Chloroplast Transformation Vector pZB1. Constructed for the introduction of heterologous genes into the spacer region between *rbcL* and *accD*, via homologous recombination. The *aadA* plastid expression cassette, which confers resistance to both spectinomycin and streptomycin, is flanked both upstream of the promoter and downstream of the terminator by multiple cloning sites.
Figure 3.14. Restriction Digest Analysis for the Construction of pHin 32-aadA. Lane 1: 1kb ladder; Lane 2: pHin 32 digested with both Sall and AflII (uncut plasmid is visualised on the agarose gel); Lane 3: pHin 32-Acpol construct digested with both Sall and AflII; Lane 4: pCP7-aadA1 digested with both Sall and AflII (releasing the 1.3kb aadA expression cassette); Lane 5: pHin 32-aadA digested with both Sall and AflII.

Figure 3.15. Restriction Digest Analysis for the Construction of pZBl. Lane 1: 1kb ladder; Lane 2: pHin 27 digested with both HindIII and EcoRI; Lane 3: pHin 32-aadA construct digested with both HindIII and EcoRI; Lane 4: pZBl digested with both HindIII and EcoRI. The pZBl digest pattern shows it to contain the 4.7kb HindIII-EcoRI fragment from pHin 32-aadA and the 980 bp fragment from pHin 27.
Figure 3.16. Restriction Analysis of pPC1. Lane 1: Circular plasmid was digested with HindIII alone, yielding a single fragment 9.1kb in length. Lane 2: “1kb ladder”. Lanes 3-7: Circular plasmid was double-digested with HindIII and one of the five restriction endonucleases whose recognition sites are located in the transformation vector Multiple Cloning Site (PmeI, KpnI, SalI, Sgfl, XbaI, see Figure 3.5). The digestion patterns, representing each one of the five cloning sites, confirm that all five are unique to the plasmid and located approximately 2kb from the HindIII site.

Figure 3.17. Restriction Analysis of pZB1. Lane 1: Circular plasmid was digested with HindIII alone, yielding a single fragment 8.3kb in length. Lane 2: “1kb ladder”. Lanes 3-5: Circular plasmid was double-digested with HindIII and one of 3 selected restriction endonucleases whose recognition sites are located in “Multiple Cloning Site 2” (see Figure 3.9; PmeI, AflII, PstI). Band patterns confirm that each site is unique to the plasmid and is located 2.25kb from the HindIII site. Lanes 6-8: Circular plasmid was double-digested with HindIII and one of 3 selected restriction endonucleases whose recognition sites are located in “Multiple Cloning Site 1” (KpnI, SalI, AvrII). Band patterns confirm that each site is unique to the plasmid and is located 0.9kb from the HindIII site.
Figure 3.18. Sequence Confirmation of the pPC1 Construct. The primers highlighted in Figure 3.4 were used to sequence the construct cloning sites, point mutation markers and polylinker insertions. Note: the sequence containing the spectinomycin resistance point mutation is excluded from this illustration because it is contained in Figure 3.9.
Figure 3.19. Sequence Confirmation of the pZBl Construct. The primers highlighted in Figure 3.12 were used to sequence all of the cloning sites involved in the assembly of the construct.
Figure 3.20. Regeneration of Tobacco Leaf Tissue bombarded with the Chloroplast Transformation Vector pPC1. **Picture A:** Tobacco leaf-disks bombarded with control (non-transformation vector) DNA and regenerated on spectinomycin free medium; **Picture B:** Tobacco leaf-disks bombarded with pPC1 DNA and regenerated on spectinomycin free medium; **Picture C:** Tobacco leaf-disks bombarded with control (non-transformation vector) DNA and regenerated on spectinomycin containing medium; **Picture D:** Tobacco leaf-disks bombarded with pPC1 DNA and regenerated on spectinomycin containing medium.
Figure 3.21. PCR Confirmation of Spectinomycin Resistance Mutation in Tobacco introduced by pPC1 Transformation Vector. Purified chloroplast DNA was isolated and used as the template to amplify the 16S rRNA gene by PCR. Restriction digestion of the fragment with AatII is diagnostic of the presence or absence of the point mutation in the AatII site that confers spectinomycin resistance. Lane 1: uncut PCR fragment from PCT1 regenerated tobacco plant; Lane 2: 1kb marker ladder; Lane 3: wild type tobacco PCR fragment digested with AatII; Lane 4: PCT1 PCR fragment digested with AatII; Lane 5: PCT2 PCR fragment digested with AatII; Lane 6: PCT3 PCR fragment digested with AatII.
Figure 3.22. Sequence Analysis of the 16S rRNA PCR Fragment from the PCT1 Tobacco Clone. The PCT1 sequence, which included the spectinomycin resistance point mutation, was obtained using the ‘138857F’ sequencing primer and aligned with the homologous sequence in *N. tabacum* (Ac. Z00044). The tobacco transformant possessed a number of *Brassica*-specific bases in this region, which confirmed that the spectinomycin resistance point mutation was derived from a homologous recombination event with the pPC1 transformation vector and not as a result of a spontaneous mutation.
Figure 3.23. PCR Analysis of the transplastomic Tobacco Plants for the Introduction of the Polylinker/XbaI Region from pPC1. Lane 1: PCR fragment amplified from pPC1 plasmid construct; Lane 2: PCR fragment amplified from B. napus chloroplast DNA; Lane 3: PCR fragment amplified from wild-type tobacco chloroplast DNA; Lane 4: PCR fragment amplified from PCT1 chloroplast DNA; Lane 5: PCR fragment amplified from PCT3 chloroplast DNA; Lane 6: pPC1 PCR digested with XbaI; Lane 7: wild-type Brassica PCR digested with XbaI; Lane 8: wild-type tobacco PCR digested with XbaI; Lane 9: PCT1 PCR digested with XbaI; Lane 10: PCT3 PCR digested with XbaI; Lane 11: pPC1 PCR digested with Pmel; Lane 12: wild-type Brassica PCR digested with Pmel; Lane 13: wild-type tobacco PCR digested with Pmel; Lane 14: PCT1 PCR digested with Pmel; Lane 15: PCT3 PCR digested with Pmel.
3.4. Discussion.

3.4.1. *B. napus* Chloroplast Transformation Vectors.

Two different plastid transformation vectors were constructed using cloned *B. napus* chloroplast DNA fragments. pPC1 was constructed to facilitate the insertion of foreign sequences into the *trnV-rps12* (3') intergenic region using “binding-type” antibiotic resistance mutations as the selection system. In contrast, pZBl uses the antibiotic detoxifying *aadA* gene to select for insertion into the *rbcL-accD* intergenic sequence. Both plasmid constructs were verified by both restriction digestion and sequence analysis.

3.4.2. Tobacco Chloroplast Transformation using pPC1.

Because a successful technique for transformation of the *B. napus* chloroplast genome and for regeneration of homoplasmic shoots has yet to be developed, the functionality of the *B. napus*-specific vector pPC1 was tested by transformation in a tobacco system. The relatively high sequence conservation that exists between the plastid genes of higher plants (see Chapter 2) has suggested the possibility of developing “universal” vectors suitable for chloroplast transformation in diverse species (Kavanagh *et al.*, 1999). Sequence conservation is particularly high for the various ribosomal genes located in the IR region, with identities between *B. napus* and tobacco being 99.5% for the 16S rRNA gene and 98.4% for the *rps12* (3') coding sequence. The *trnV-rps12* (3') spacer region (and flanking sequences) has previously been transformed in tobacco using constructs containing either *S. nigrum* (Kavanagh *et al.*, 1999) or petunia (DeGray *et al.*, 2001) chloroplast DNA sequences. However, all of these sequences are from relatively closely related species within the same family, the Solanaceae.

3.4.2.1. Transformation Efficiency of pPC1 in Tobacco.

A total of ten tobacco leaves were bombarded with pPC1-coated tungsten particles. Three independent green shoots were regenerated under spectinomycin selection conditions. PCR analysis revealed two of the clones to have incorporated the spectinomycin resistance marker from the *Brassica*-specific transformation
vector, whereas one was a selection system escape. This efficiency is comparable to the results obtained by Svab et al., (1990), who recovered 56 spectinomycin resistant shoots from 150 bombarded leaves. Approximately one third of these proved to be spontaneous mutants, independent of transformation vector bombardment.

The two pPC1 transplastomic tobacco plants were then screened for streptomycin resistance by transferring the green tissue into streptomycin containing medium. Both of the propagated clones bleached and neither developed roots, indicating sensitivity to the antibiotic. PCR analysis performed on the \textit{trnV-rps12} (3') intergenic sequence also showed that neither of the spectinomycin resistant transformants had incorporated the transformation vector multiple cloning site. These results naturally lead to the conclusion that the homologous recombination events that caused the introduction of spectinomycin resistance marker was localised to a small region within the 16S rRNA gene and did not include either the \textit{rps12} (3') or the intergenic regions. The 99.5% sequence similarity that exists between the 16S rRNA genes of the \textit{Brassica} and tobacco plastid genomes is clearly sufficiently homologous to allow recombination to occur within the coding sequence. Whether or not the 98.4% identity value for \textit{rps12} (3') is sufficient to allow efficient recombination can only be answered by performing numerous additional bombardments. However, it is unlikely that homologous recombination events would occur between the highly divergent intergenic regions.

### 3.4.2.2. Universal Transformation Vectors.

Although there have been numerous cases of plastid sequences from one higher plant species being used as the targeting sequence to transform the genome of another, it is not yet fully understood to what degree sequence divergence affects the efficiency and extent of homologous recombination in higher plant chloroplasts. Kavanagh et al. (1999) obtained a high tobacco transformation efficiency using the moderately divergent (2.4% average divergence over 7.8kb) \textit{Solanum} plastid sequence vector (pSSH1), similar to the frequency obtained using completely homologous tobacco sequences. However, DeGray et al. (2001) obtained a reduced transformation frequency when the vector flanking sequence
was derived from the petunia plastome. They discovered that only 7% of bombarded tobacco leaves yielded a transgenic plant for the petunia-specific vector compared to 91% when completely homologous sequence was used. It must be noted that the two former experiments used different transformation parameters, with Kavanagh et al. (1999) using a PEG-mediated delivery and a "binding-type" selection system, whereas DeGray et al. (2001) utilised biolistic bombardment to introduce a chimeric aadA expression cassette.

In order to fully ascertain the efficiency with which pPCI can transform tobacco, transformed lines that contain B. napus intergenic cloning site/heterologous gene and not just an antibiotic resistance marker will have to be obtained. Since transformation frequencies are influenced by sequence divergence, it can be expected that a large number of leaves will have to be bombarded to obtain such a transformation event. Since tobacco, petunia and potato all belong to the Solanaceae (in the Asterid clade), whereas B. napus belongs to the Brassicaceae (in the Rosid clade) (Angiosperm Phylogeny Group, 1998) the flanking (or targeting) sequences in pPCI are considerably more divergent relative to the tobacco chloroplast genome than the flanking sequences in the S. nigrom-derived transformation vector pSSH1 (discussed above).

Although the "binding-type" selection system has been used to obtain plastid transformation frequencies comparable to those derived by aadA-based selection, it has only been used in PEG-mediated transformation experiments (Kavanagh et al., 1999). The fact remains that the use of the dominant spectinomycin resistance marker gene (i.e. aadA) yields a much greater frequency for biolistics (Svab and Maliga, 1993). Therefore, in light of the possibility of sequence divergence reducing plastid transformation frequency, it might be wise to determine whether biolistic or PEG-mediated delivery of the pPCI vector is more efficient in producing tobacco transformants. Such experiments will clarify the efficiency of using the B. napus-specific construct for tobacco plastid transformation and its potential for use as a "universal" vector. Alternatively, an aadA expression cassette could be placed into the trnV-rps12 (3') intergenic region
of pPC1, which might increase the efficiency of transformation to the levels obtained following leaf bombardment.

3.4.3. The *B. napus*-specific vector pZB1.

The pZB1 transformation vector was not tested in plastid transformation experiments in tobacco because the flanking sequences were based on the *B. napus* *rbcL* and *accD* genes which were found to be considerably more divergent between the two species than the sequences in pPC1 (see Tables 2.5 and Figure 2.17). Although it may not be suitable for use as a "universal" vector it should, like pPC1, prove to be valuable for the manipulation of the *B. napus* plastid genome once an efficient transformation and regeneration protocol has been developed for *Brassica*. 
CHAPTER 4
A CHLOROPLAST-TARGETTED CRE RECOMBINASE.

4.1. Introduction.

4.1.1. Site Specific Recombination Systems.

In contrast to homologous recombination which requires extensive nucleotide sequence homology and numerous protein factors, site-specific recombination occurs between short (approximately 35 bp) highly conserved sequence elements (recombination sites) and requires only a single recombinase protein. Several site-specific recombination systems have been characterised in detail in yeast, bacteria and phage and used to direct the integration, excision or inversion of transgenes in heterologous organisms including plants. Three such systems are of particular interest because of the simplicity of their requirements and the base-pair precision of the recombination reaction. These are the Cre/loxP, FLP/frt and R/rs systems and are discussed below.

4.1.1.1. The Cre Recombinase/loxP system.

The 38.5kDa Cre recombinase encoded by the E. coli bacteriophage P1 (Sternberg and Hamilton, 1981) is the only protein factor necessary to catalyse recombination between target loxP sites in DNA (Sauer and Henderson, 1989). The recombinase recognition site loxP consists of a 34 bp sequence containing two 13 bp inverted repeats separated by an 8 bp unidirectional spacer region (Hoess et al., 1982). Each of the 13 bp elements binds a Cre recombinase molecule and these function in tandem to direct recombination (Mack et al., 1992). The system serves two functions in the lysogenic life cycle of the phage: firstly, it allows circularization of the viral DNA so it can enter the host (Segev and Cohen, 1981; Sternberg et al., 1981); secondly, it breaks down dimer molecules that may form as a result of DNA replication (Austin et al., 1981).
Purified Cre recombinase can recombine loxP sequences in vitro with only Mg²⁺ or spermidine needed as a co-factor (Abremski and Hoess, 1984). The kinetics of the reaction are fast, with maximal recombination occurring within 2-3 minutes at 37°C. Results from this study also demonstrate that Cre acts stoichiometrically, with both enzyme and DNA concentrations being critical factors determining the recombination frequency. The enzyme uses no energy and recombination reactions are freely reversible resulting in equilibrium (Abremski and Hoess, 1984). Purification of bacterially produced Cre recombinase for in vitro studies has been facilitated by its expression as a fusion protein with the E. coli maltose binding protein (MBP; Kolb and Siddell, 1996). Subsequent cleavage of the fusion protein was found not to be necessary, as the biochemical activity of MBP-Cre was virtually identical to that of the native Cre.

Cre recombinase can interact with loxP sites on both circular and linear double-stranded DNA molecules (reviewed in Kilby et al., 1993). The type of recombination event depends on the number and relative orientation of the loxP sites. Recombination between a pair of loxP sites in the same orientation results in excision of the intervening DNA. Thus if a transgene in flanked by loxP sites having the same orientation it will be excised by Cre recombinase. The reverse reaction (e.g. between two different DNA molecules each possessing a single loxP site) will cause the integration of one molecule into the other. A DNA fragment flanked by loxP sites that are in opposite orientations will be inverted by Cre recombinase activity. Also loxP sites on separate linear molecules cause mutual exchange of DNA regions following recombination (Kilby et al., 1993).

4.1.1.2. Other Recombinase Systems.

FLP recombinase is responsible for catalysing recombination at frt sites (Andrews et al., 1985). This 43kDa protein is encoded by the 2μ plasmid of the yeast Saccharomyces cerevisiae. Frt sites are identical in size and structure, but differ in sequence, to loxP. The repertoire of possible recombination events for the FLP/frt system is the same as that of Cre/loxP system (Kilby et al., 1993). Finally, the R/rs recombination system found on the pSR1 plasmid of Zygosaccharomyces


*roxi*, is based on the interaction of 31 bp *rs* recombination elements with the so-called R recombinase (Araki *et al.*, 1985).

### 4.1.2. Heterologous Expression of Recombinases in Plants.

Expression of the *cre* gene has been extensively used to direct site-specific recombination in various eukaryotic systems, including mammalian cell lines (Sauer and Henderson, 1988; 1989; 1990; Fukushige and Sauer, 1992) and in transgenic plants (Odel *et al.*, 1990). The FLP/frt and R/rs recombination systems have also been shown to function in plants (Kilby *et al.*, 1995; Onouchi *et al.*, 1995).

#### 4.1.2.1. Site-specific Gene Insertion using the Cre/loxP System.

Site-specific integration of DNA using Cre recombinase has the drawback that Cre-mediated recombination events are reversible and kinetically favour excision. Albert *et al.* (1995) reported the first incidence of efficient Cre-mediated integration of a transgene into a *loxP* site that had previously been placed in the nuclear genome of tobacco. Two strategies were utilised to limit post-integration Cre activity. The first of these involved promoter displacement using the transgene to interrupt Cre expression. Integration was assisted by the use of mutant *loxP* sequences that favour the forward over the reverse direction of recombination. The second strategy was to provide transient expression of the recombinase, with the *cre* gene cassette not being stably incorporated into the plant genome. Successful transgene integrations were recovered using both strategies and were shown to have occurred without additional insertions elsewhere in the genome.

Recently the highly efficient *Agrobacterium* gene transfer system was used in conjunction with Cre recombinase to perform site-specific integration of T-DNA into the nuclear genome of *A. thaliana* (Vergunst and Hooykaas, 1998). Such Cre/loxP-mediated insertions required conversion of the linear single-stranded form of the transgene into a circular double-stranded form. Circularisation was made possible by flanking the T-DNA with two directly oriented *loxP* sites.
4.1.2.2. Site-Specific DNA Excision from the Nuclear Genome.

There have been numerous reports of site-specific excisions from the nuclear genome of plants using the Cre/lox recombination system (Bayley et al., 1992; Dale and Ow, 1990; Odel et al., 1990). One of the more attractive applications of such a technique is for the precise removal of a selectable marker gene following transformation (Dale and Ow, 1991). Recombination between two loxP sites flanking an antibiotic resistance gene cassette was catalysed by the introduction of Cre recombinase by cross-fertilisation. The use of a conditional lethal dominant gene in conjunction with transient Cre expression allowed the production of selectable marker-free transgenic plants without the need for sexual crosses (Gleave et al., 1999). The strategy combined a kanamycin resistance gene (npt II) and the cytosine deaminase gene (codA) in a single cassette (npt II-codA) flanked by loxP sites in the same orientation. Cells containing the codA gene will be killed if the growth medium contains 5-fluorocytosine (5-fc). However if the npt II-codA cassette is excised by transient expression of Cre recombinase, shoots that are insensitive to 5-fc and kanamycin sensitive can be regenerated. This was shown to be due to Cre-mediated excision of the loxP-flanked cassette.

In an attempt to further advance the efficiency of DNA excision, a system was developed in Arabidopsis whereby expression of Cre recombinase was tightly controlled by a steroid-inducible promoter (Zou et al., 2001). The complete removal of the marker gene was demonstrated upon β-estradiol administration in both germ line and somatic cells.

4.1.3. Developing a Cre/loxP Recombination System for Chloroplasts.

This Chapter describes experiments aimed at developing a marker gene excision system for chloroplasts based on Cre/loxP, which would facilitate the removal of the aadA selectable marker gene after selection for homoplasmic transplastomic lines had been achieved.
The experimental aims were two-fold:
(a) to develop a plastid transformation construct that would facilitate Cre-mediated marker gene excision and simultaneously provide visual confirmation that excision had occurred.
(b) to develop a means of delivering Cre recombinase into chloroplasts.

The strategy chosen for visual confirmation of marker gene excision was based on the activation of green fluorescent protein (GFP) expression as a direct result of Cre-mediated excision. This involved the insertion of an antibiotic resistance gene excision cassette (loxP-aadA-loxP) into the GFP gene. The presence of the marker gene cassette within the GFP gene prevents its expression and consequently cells containing this construct will not show green fluorescence under UV light. However, exposure of the construct to Cre recombinase restores the GFP open reading frame (orf) by catalysing excision of the resistance gene cassette and cells in which this has occurred will show green fluorescence.

The strategy chosen for delivering Cre recombinase into chloroplasts was based on the construction of a gene fusion comprising the transit peptide sequence of the Rubisco small subunit (SSU-TP) gene fused with Cre recombinase. SSU transit peptide sequences have previously been shown to efficiently direct the import of fusion proteins into chloroplasts (Wong et al., 1992).
4.2. Materials and Methods.

4.2.1. Materials.

4.2.1.1. Plasmids.

pBAD-GFP (Crameri et al., 1996), which contains the GFP gene under the control of the arabinose-inducible BAD promoter was used in this Chapter to generate the various pBAD-lox-GFP derivatives used for assaying Cre recombinase activity.

pGST-Cre (unpublished plasmid; provided by Dr. Mark Harper) was the source of the cre gene used in the construction of both pRokCre and pRok8Cre, the Agrobacterium vectors used to transform the nuclear genome of N. tabacum.

pCAR2 (unpublished plasmid) was the source of the RUBISCO SSU gene (rbcS) encoding the transit peptide sequences introduced in front of the cre gene in pRok-Cre to direct the cytoplasmically translated Cre recombinase to the chloroplast.

pRok10 (unpublished plasmid) was the pBin19-based vector used in the construction of pRok-Cre.

pRok8 (unpublished plasmid) was the pBin19-based vector used in the construction of pRok8-Cre.

4.2.1.2. Antibodies.

Anti-Cre rabbit polyclonal antibody (Novagen) was used to detect Cre protein during western blot analysis.

Anti-Rabbit IgG Alkaline Phosphatase Conjugate (Novagen) was the secondary antibody used to identify Anti-Cre antibody bound to the nitrocellulose membrane.

Anti-GFP rabbit polyclonal antibody (Invitrogen©) was used to detect GFP protein during western blot analysis.

Anti-Rabbit IgG Alkaline Phosphatase Conjugate (Sigma Immuno Chemicals), developed in goat, was the secondary antibody used to identify anti-GFP antibody bound to the nitrocellulose membrane.
4.2.1.3. Oligonucleotides used during the Assembly of the Target Construct for the Cre Recombinase Activity Assay.

The following oligonucleotides were annealed to form the synthetic DNA fragment that was inserted into the \textit{SalI-AvrII} restriction sites of pCP-aadA\textsubscript{1}, introducing a 34 bp \textit{loxP} recombination site 5' of the \textit{aadA} cassette promoter (\textit{Prrn}):

\begin{verbatim}
loxP-aadAf: 5'-CTA GGT ATA ACT TCG TAT AAT GTA TGC TAT ACG AAG TTA TTT AAT TAA GAT CTA G-3'.
loxP-aadAr: 5'-TCG ACT AGA TCT TAA TTA AAT AAC TTC GTA TAG CAT ACA TTA TAC GAA GTT ATA C-3'.
\end{verbatim}

The following oligonucleotides were annealed to form the synthetic DNA fragment that was inserted into the unique \textit{NheI} restriction site of pBAD-GFP, introducing a 34 bp \textit{loxP} recombination into the \textit{GFP} orf. Although the \textit{loxP} sequence was translated as part of the \textit{GFP} gene (adding 21 extra amino acids to the N-terminus), the oligonucleotide design ensured the correct reading frame was maintained:

\begin{verbatim}
lox-GFPf: 5'-CTA GCT TAA TAA CTT CGT ATA GCA TAC ATT ATA CGA AGT TAT CGT TTA AAC AAG GCC TAG GAT-3'.
lox-GFPr: 5'-CTA GAT CCT AGG CCT TGT TTA AAC GAT AAC TTC GTA TAA TGT ATG CTA TAC GAA GTT ATT AAG-3'.
\end{verbatim}

The following primers were used to introduce an alternative translation start-codon (methionine residue) downstream of the \textit{loxP} site in the pBAD-lox-GFP construct. The primers yielded a PCR product (approximately 280 bp) that was digested with both \textit{AvrII} and \textit{NcoI} and cloned into pBAD-lox-GFP, replacing the existing \textit{AvrII-NcoI} fragment (172 bp):

\begin{verbatim}
GFPiMet: 5'-CAA GGC CTA GGA TAT GGC TAG CAA AGG AGA AG-3'.
GFPr2: 5'-TCA TGC CGT TTC ATA TGA TCC GGA TAA CG-3'.
\end{verbatim}

Three primers were used to introduce additional "linker" amino acid codons into pBAD-lox-GFP in order to create a greater distance between the co-translated
loxP site and the GFP domain. The primer GFP-GM introduced a single additional glycine-encoding codon immediately upstream of a supplementary GFP start-codon, whereas GFP-GG introduced two glycines. GFP-GGAG was used to insert a glycine-glycine-alanine-glycine linker in between the translated loxP and GFP domains. This primer resulted in a construct that did not contain an additional translation start-codon downstream of the loxP site, so that translation would have to begin upstream of the loxP domain. All three linkers were introduced via PCR using GFPPr2 as the reverse primer. Each PCR product (approximately 280 bp) was digested with both AvrII and Ncol and cloned into pBAD-lox-GFP, replacing the existing AvrII-Ncol fragment (172 bp):

- GFP-GM: 5'-CAA GGC CTA GGT ATG GCT AGC AAA GG-3'.
- GFP-GG: 5'-CAA GGC CTA GGT GGT ATG GCT AGC AAA GG-3'.
- GFP-GGAG: 5'-CAA GGC CTA GGT GGT GCT GGC GCT AGC AAA GGA G-3'.

The following pair of oligonucleotides were annealed and cloned into the NheI restriction site located in pBAD-lox-GFP, resulting in derivatives containing additional nucleotides (continuous with the existing reading frame) immediately downstream of the cassette's unique translation start codon. It was hoped that this linker would destabilise any hairpin structures in the lox-GFP mRNA transcript, resulting from the inverted repeat sequences in loxP, which might be affecting the efficiency of translation.

- GFP-ins1a: 5'-CTA GCG GAG GCT CTG GCT TGA GTG GAG-3'.
- GFP-ins1b: 5'-CTA GCT CCA CTC AAG CCA GAG CCT CCG-3'.

A similar oligonucleotide pair was also synthesised to form a set of linker fragments containing a random nucleotide base at five different residue positions. The annealed oligonucleotides were cloned into the NheI restriction site in pBAD-lox-GFP:

- GFP-insNa: 5'-CTA GCG GAG GCT C (AGCT) (AGCT) (AGCT) (AGCT) (AGCT) TGA GTG GAG-3'.
- GFP-insNb: 5'-CTA GCT CCA CTC A (AGCT) (AGCT) (AGCT) (AGCT) (AGCT) GAG CCT CCG-3'.

102
4.2.1.4. Sequencing Primers used to confirm the Correct Assembly of pBAD-lox-aadA-N-GFP.

The following primer was designed so that, along with GFFr2, the cloning junctions utilised during the construct's assembly can be sequenced:

BADf: 5'-'GAT TAG CGG ATC CTA CCT GAC GC-3'.
Prrm-Rev: 5'-GAA ATA TAG CCA TCC CTG CCC-3'.

4.2.1.5. Oligonucleotides used during the Assembly of the Cre Recombinase Nuclear Expression Cassettes pRok-Cre and pRok8-Cre.

The following oligonucleotides were annealed to form the polylinker that facilitated cloning of the cre gene and construction of pRok-Cre and pRok8-Cre:

Jul1: 5'-AGC TTC TAG ACA ATT GGC CAT GGA CGT CGA ATT CAC TAG TGA GCT CC-3'.
Jul2: 5'-AAT TGG AGC TCA CTA GTG AAT TCG ACG TCC ATG GCC AAT TGT CTA GA-3'.

4.2.1.6. PCR Primers used to Screen for Transgenic Plants Transformed with pRok-Cre and pRok8-Cre.

CarSSUlf: 5'-GAG CAA TGT TGC TCA AGC TAA CAT G-3'.
CreF3: 5'-CGC GAT TAT CTT CTA TAT CTT CAG GC-3'.
CreR3: 5'-TAT CTC TGA CCA GAG TCA TCC TTA G-3'.

4.2.1.7. Sequencing Primers used to confirm the Correct Assembly of pRok-Cre and pRok8-Cre.

In addition to CarSSUlf, the following primer was designed to facilitate the sequencing of the cloning junctions utilised during construct assembly:

CreR2: 5'-GGA CAG AAG CAT TTT CCA GGT ATG CTC AG-3'.

103
4.2.2. Methods.

4.2.2.1. Preparation of Electroporation Competent *Agrobacterium* Cells.

The *Agrobacterium* strain LBA4404 was streaked on LB medium containing 500μg/ml streptomycin and incubated at 28°C for 48 hours. A single colony was used to inoculate 10ml YEB medium (0.5% peptone, 0.1% yeast extract, 0.5% beef extract, 0.5% sucrose, 2mM MgSO₄) containing 500μg/ml streptomycin and shaken overnight at 250rpm at 28°C. A 1ml aliquot of this subculture was then used to inoculate 100ml fresh YEB medium (containing streptomycin), which was then shaken overnight at 250rpm at 28°C. Cells were grown until the culture reached an OD₆₀₀ = 0.5, i.e. early- to mid-log phase. The cells were harvested by centrifuging for 15 minutes at 4000 x g at 4°C. The cells were then suspended in 10ml of cold 1mM HEPES pH7 and centrifuged as above. The cells were then suspended in 10ml of cold 1mM HEPES pH7, 10% glycerol and again centrifuged as above. The cells were washed once more with 5ml of cold 1mM HEPES pH7, 10% glycerol before finally being gently suspended in 0.5ml of cold 1mM HEPES pH7, 10% glycerol and divided into 60μl aliquots.

4.2.2.2. *Agrobacterium* Transformation by Electroporation.

*Agrobacterium* electroporation was performed in accordance with the recommendations of the protocol manual from Life Technologies® (GIBCO BRL®). A 20μl aliquot (for one electroporation) of *A. tumefaciens* competent cells was thawed on ice. 50ng of the purified DNA construct was gently mixed with the cells and incubated on ice for 5 minutes. The Cell-Porator® (Life Technologies®) was set for *Agrobacterium* electroporation: 2.4kV; 330μF; 4kΩ. The cells were transferred to a sterile cuvette with a 1mm gap that was placed into the electroporation chamber. Following a 6ms pulse, the cells were transferred to a vial containing 1ml of room temperature YEB medium, which was then shaken for 3 hours at 225rpm. The cells were spread on YEB agar containing 500μg/ml streptomycin and 50μg/ml Kanamycin (to select for pRok-Cre and pRok8-Cre transformants) and incubated at 28°C for 48 hours.
4.2.2.3. Transformation of Tobacco.

*Agrobacterium tumefaciens* mediated transformation of tobacco nuclear DNA was performed by the method described by Horsch *et al.* (1985). *Agrobacterium tumefaciens* containing the various cre gene constructs was grown overnight in LB broth in the presence of streptomycin 500mg/L and kanamycin 50mg/L. Cells were pelleted by centrifugation at 4000 rpm for 10 min at 4°C and resuspended in MS liquid medium (Murashige and Skoog, 1962). The suspension was incubated for 30 minutes in a sterile environment with freshly prepared leaf discs from *in vitro* tobacco plants. The leaf discs were then transferred onto shoot regeneration NBM medium (MS containing 0.8% agar, benzylaminopurine (BAP) 1μg/ml, α-naphthalene acetic acid (NAA) 0.1μg/ml and B5 vitamins; inositol 100μg/ml, thiamine 10μg/ml, nicotinic acid 1μg/ml and pyridoxine-HCl 1μg/ml). After two days of cocultivation the leaf discs were transferred to NBM medium containing carbenicillin 200mg/L and kanamycin 50mg/L. Individual shoots that appeared on selective medium after 3-4 weeks were cut and transferred to MS medium containing the same selection. Shoots that rooted in the presence of selection were planted in the greenhouse for analysis.

4.2.2.4. Preparation of Tobacco DNA Extracts.

Leaf tissue was homogenized in Extraction Buffer (2% SDS, 50mM NaCl, 50mM EDTA, 100mM Tris, 10mM β-mercaptoethanol and 100mg/ml Proteinase K) using a pestle and mortar. The homogenate was then incubated at 85°C for 5 minutes, placed in ice for 5 min, returned to 85°C for a further 5 min and replaced in ice for 5 min. The mixture was briefly vortexed before being centrifuged in a microfuge for 5 minutes. The supernatant was recovered and the solid debris was discarded. The supernatant was then subjected to 3 phenol extractions followed by spin-dialysis through a TE equilibrated Sepharose CL6B mini column to remove salts and residual phenol (Murphy and Kavanagh, 1988).

4.2.2.5. Preparation of Tobacco Protein Extracts.

Total protein extracts for western blots were prepared by homogenizing leaf tissue in Protein Extraction Buffer (60mM Tris base, pH 6.8, 5mM EDTA, 1mM PMSF, 10mM β-mercaptoethanol, 0.1% Triton X-100) using a mortar and pestle on
ice and a 2:1 buffer:tissue ratio. The homogenate was centrifuged briefly to remove insoluble debris.

Chloroplasts were prepared by gently grinding approximately 1g of leaf tissue in ice-cold extraction buffer containing 330mM Sorbitol, 50mM Tris-HCl, pH 8.0, 2mM MgCl₂. This was then filtered through 2 layers of cheesecloth. The filtrate was centrifuged for 10 minutes at 500g in a Sorvall RC-5B centrifuge using a prechilled GSA rotor to pellet the cell nuclei. The supernatant was decanted into a fresh tube and centrifuged for 10 minutes at 3000g (4°C) to pellet chloroplasts. This pellet was then resuspended in 100μl deionised water to lyse the chloroplasts.

4.2.2.6. Western Blotting.

SDS-PAGE and western blotting was carried out essentially as described by Sambrook et al. (1989). Total leaf, chloroplast and bacterial protein extracts were prepared for SDS-PAGE as described by Laemmli (1970). Protein extracts were resolved in 8-10% separating gels in the BioRAD SDS-PAGE apparatus run at 140 volts until the dye front eluted into the running buffer. Protein gels were prepared and stained with GELCODE® Blue Stain Reagent. Duplicate protein extracts for western blotting were resolved in polyacrylamide gels at the same time and were transferred to nitrocellulose membranes (Schleicher & Schuell) in a BioRad Transblot™ apparatus at 150mA overnight. The membrane was immersed in a blocking solution (10mM Tris-HCL, pH7.5, 350mM NaCl, 0.5% Tween-20 (polyoxyethylene-sorbitan monolaurate) and 2-9% nonfat dried milk) for 1 hour to prevent non-specific binding of immunoglobulins. The membrane was then incubated with an appropriate dilution (in accordance to the recommendations of the supplier) of the primary antibody in Buffer A (10mM Tris-HCL, pH7.5, 150mM NaCl, 0.5% Tween-20). Alkaline phosphatase coupled anti-rabbit antibodies were used in the secondary immunological reaction and the localisation of alkaline phosphatase was visualised by treatment with 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium.
4.2.2.7. *In vitro* Cre recombinase Activity Assays.

*In vitro* activity assays using purified Cre recombinase (Novagen) and transgenic plant protein extracts were carried out in accordance with the recommendations of Abremski and Hoess (1984). Both the target assay plasmid (pBAD-lox-aadA-N-GFP) and the enzyme concentrations were optimised for each of the Cre recombinase sources: (a) the commercial enzyme (b) total leaf protein extracts and (c) chloroplast protein extracts.

In the case of the total leaf protein Cre recombinase assay, fresh leaf tissue was homogenised (using a pestle and mortar) in Cre activity buffer (50mM Tris-Hcl, pH7.5, 33mM NaCl, 10mM MgCl₂, 500μg/ml BSA; Abremski and Hoess, 1984), supplemented with 0.01% Triton-X to lyse the chloroplasts. The homogenate was centrifuged for 3-5 seconds in a microfuge to sediment the cellular debris. The supernatant was recovered and its Cre recombinase activity assayed by incubation with the assay plasmid for 15 minutes at 37°C. The mixture was purified by a spin-dialysis treatment on a Sepharose CL6B spin-dialysis column equilibrated with TE (Murphy and Kavanagh, 1988). The extent of Cre-mediated recombination of the assay plasmid was determined by transformation into *E. coli*. The transformation mixture was plated on LB-agar containing 100μg/ml ampicillin and 0.2% arabinose. Following a 48-hour incubation at 37°C the numbers of non-fluorescent colonies (containing non-recombined assay plasmid) and fluorescent colonies (containing recombined plasmids) were determined under long-wave (366nm) UV-light.

For the chloroplast Cre recombinase assay, chloroplast purification and lysis was carried out as described above (Section 4.2.2.8). The chloroplast lysate was adjusted to the composition of Cre assay buffer and 20μl of the adjusted lysate incubated with the target plasmid for 15 minutes at 37°C. The remaining analytical steps were the same as described for the other Cre assays.
4.3. Results.


The plasmid ‘pBAD-lox-aadA-N-GFP’ was constructed to facilitate the detection of Cre recombinase activity by the activation of GFP expression. A physical map of the final construct is detailed in Figure 4.3. The non-recombined version of plasmid contains an aadA resistance gene cassette flanked by loxP sites. This cassette is located within the GFP open reading frame. Exposure to Cre recombinase causes excision of the aadA cassette and restoration of the GFP reading frame despite the fact that a single loxP site remains. Ultimately, ‘pBAD-lox-aadA-N-GFP’ will be incorporated as part of a chloroplast transformation vector to allow in planta visualisation of marker gene removal following delivery of nuclear-expressed Cre to the plastids. Because both genes (GFP and aadA) were also functional in E. coli, the plasmid was used throughout this chapter as a target substrate for the assay of Cre recombinase activity produced in transgenic tobacco plants.

The following sections describe the numerous steps required to construct ‘pBAD-lox-aadA-N-GFP’:

4.3.1.1. Introduction of a loxP Element at one end of the aadA Expression Cassette in ‘pCP7-aadA1’.

The construction of the plasmid ‘aadA-loxP’ is diagrammed in Figure 4.1. The synthetic oligonucleotides ‘loxP-aadAf’ and ‘loxP-aadAr’ were annealed to produce a double-stranded loxP DNA fragment that had a SalI site at one end and an AvrII site at the other. The plasmid ‘pCP7-aadA1’, which contains unique SalI and AvrII sites immediately upstream of the Prrn element, was digested using both restriction endonucleases. A simple ligation reaction resulted in the insertion of the 55 bp loxP oligonucleotide fragment into the plasmid. The use of two different restriction sites in the cloning step meant the directionality of the loxP sequence in
the construct could be controlled. A correctly cloned ‘aadA-loxP’ construct was identified using the sequence derived from the ‘Prm-Rev’ primer.

4.3.1.2. Introduction of a loxP Element into the GFP Open Reading Frame in pBAD-GFP.

The construction of ‘pBAD-lox-GFP’ is diagrammed in Figure 4.2. The oligonucleotides ‘loxP-GFPf’ and ‘loxP-GFPr’ were annealed to produce a DNA fragment with two NheI cohesive ends, although only one of these was cleavable in the final ligated product. Digestion of ‘pBAD-GFP’ with NheI cuts the plasmid at a site immediately adjacent (3’) to the GFP start (ATG) codon. A ligation reaction involving NheI-cleaved pBAD-GFP the annealed loxP oligonucleotides resulted in the insertion of the 63 bp loxP fragment into the GFP gene. Prospective clones were sequenced using the ‘BADf’ primer to ensure the construct (‘pBAD-lox-GFP’) possessed a monomeric insertion in the correct orientation. In this orientation, the inserted loxP site is flanked by a unique NheI site close to the ATG of GFP (i.e. nearest to the P_{BAD} promoter) and a polylinker cloning site (P_{mel}-StuI-AvrII; required for subsequent cloning steps) at the other side. Such an orientation was critical for the functioning of the loxP site in the final construct.

Furthermore, although the loxP element was inserted into the GFP gene, it does not disrupt the reading frame (i.e. the introduced loxP fragment does not contain stop codons or cause frame-shifts). Consequently expression of lox-GFP from the pBAD promoter should produce a GFP fusion protein which contains an extra 21 amino acids (due to loxP) at its N-terminus and which we predicted would retain green fluorescence under UV light.

In order to determine whether the lox-GFP fusion protein retained the same fluorescence intensity as wild-type GFP, ‘pBAD-lox-GFP’ was transformed into E. coli DH5α and grown on LB-agar containing 0.2% arabinose, which was required for promoter induction. Following an extended incubation period of 48 hours at 37°C, the culture was illuminated using 366nm UV-light. Unfortunately the resulting fluorescence was considerably diminished compared to cultures
transformed with the parent vector ‘pBAD-GFP’. It was unclear at this point whether the introduction of the 63 bp $\text{loxP}$ fragment into the $\text{GFP}$ orf was inhibiting protein translation or if the extra 21 amino acids encoded by the $\text{loxP}$ insert had caused a structural change to the protein that altered its ability to fluoresce. Therefore, several different ‘pBAD-lox-GFP’ derivative constructs were generated in order to attempt to identify novel $\text{loxP}$ insertions that might give improved green fluorescence levels and to explain the observed reduction in fluorescence intensity. These derivative constructs are illustrated in Figure 4.3.

4.3.1.2.1. **Construction of Derivatives of ‘pBAD-lox-GFP’ for Improved GFP Expression.**

The construct ‘pBAD-lox-GFP-iMet’ was designed to introduce an alternative $\text{GFP}$ translation initiation codon downstream of the $\text{loxP}$ sequence in ‘pBAD-lox-GFP’. The mutation was introduced into the parent construct via the primer ‘GFPiMet’ in a PCR reaction that had ‘GFPr2’ acting as the complementary primer. This PCR reaction (95°C for 30 sec; 55°C for 30 sec; 72°C for 30 sec; through 25 cycles) yielded a product approximately 280 bp in length. This fragment was subsequently digested with both $\text{AvrII}$ and $\text{NcoI}$ and cloned into ‘pBAD-lox-GFP’, replacing the existing $\text{AvrII-NcoI}$ fragment (172 bp) that was removed from the vector digest by gel purification. Figure 4.4 shows an agarose gel containing the PCR product and vector digest generated during the construction of ‘pBAD-lox-GFP-iMet’. The plasmid ‘pBAD-lox-GFP-iMet’ was transformed into $\text{E. coli DH}α$ and grown on LB-agar containing 0.2% arabinose for 48 hours at 37°C. Illumination under UV-light showed that colonies did not fluoresce any more brightly than those of the original ‘pBAD-lox-GFP’.

The effect of different $\text{loxP}$ derivatives on GFP expression was investigated further by the creation of three constructs that possessed additional neutral amino acids separating the translated $\text{loxP}$ site and the native GFP domains. ‘pBAD-lox-GFP-GM’ contains one extra glycine codon in front of a supplementary $\text{GFP}$ start codon, downstream of the $\text{loxP}$ sequence. ‘pBAD-lox-GFP-GG’ has two extra glycine residues in the $\text{loxP-GFP}$ translatory product as well as the supplementary start codon. The protein product of ‘pBAD-lox-GFP-GGAG’ contains four
additional amino acids (glycine-glycine-alanine-glycine) separating the loxP and GFP domains. ‘pBAD-lox-GFP-GGAG’ does not have the supplementary start codon, so only GFP fused to the loxP domain can be expressed. All three constructs were assembled using a PCR strategy similar to that described for ‘pBAD-lox-GFP-iMet’, in order to introduce the additional codons. Each modified plasmid was transformed into E. coli and observed for UV-fluorescence intensity as described. The results showed that the three derivatives fluoresced no more brightly than cultures expressing the original ‘pBAD-lox-GFP’ construct.

The unchanged levels of fluorescence observed for the various 'pBAD-lox-GFP' derivatives described above led us to conclude that insertion of the loxP element into GFP might be affecting the efficiency of GFP translation. This conclusion was based on the observation of Liu et al. (1998) that RNA transcripts of the loxP element can potentially fold into a 13 bp stem-loop structure. This type of secondary structure in mRNA can strongly inhibit translation. One approach to solving the problem is to either modify or select for loxP derivatives that reduce the stability of the stem-loop structures. To this end, a further set of 'pBAD-lox-GFP' derivatives was constructed.

The plasmid ‘pBAD-lox-GFP-ins1’ was constructed by inserting a 27 bp DNA fragment, formed by annealing the oligonucleotides ‘GFP-ins1a’ and ‘GFP-ins1b’, into the unique NheI site (upstream of loxP) of ‘pBAD-lox-GFP’. The vector digest was phosphatased and the artificially synthesised fragment was kinased, so that under ligating conditions the fragment was capable of inserting in two possible orientations either as a monomer or as numerous different multimer combinations. The ligation mixture was transformed into E. coli, spread on LB-agar containing 0.2% arabinose and incubated for 48 hours at 37°C. Colonies were screened for GFP fluorescence using 366nm UV-light. Eight clones were selected and streaked on LB-agar containing 0.2% arabinose and following a ‘miniprep’ the plasmid in each clone was sequenced using the primer ‘GFPr2’ (data not shown). The clone that exhibited the brightest fluorescence, ‘pBAD-lox-GFP-ins1-6’ (i.e. clone number 6), contained a plasmid with a monomeric insertion of the ‘GFP-ins1’ fragment. Its fluorescence was noticeably brighter than that of the culture
transformed with ‘pBAD-lox-GFP’. This suggests that the insertion may have relaxed the mRNA stem-loop structure at the loxP inverted repeat sequence so that translational efficiency was improved.

In a final attempt to further improve GFP fluorescence, a set of random oligonucleotide fragments, created by annealing the oligonucleotides ‘GFP-insNa’ and ‘GFP-insNb’, were cloned into the Nhel site of ‘pBAD-lox-GFP’. The 27 bp fragments, with Nhel cohesive overhangs, were kinased and inserted into the phosphatased vector digest by ligation. Having transformed the ligation mixture into E. coli, the colonies were screened for UV-fluorescence as described previously. One particular clone, that containing the construct ‘pBAD-lox-GFP-insN-8’ (i.e. clone number 8), displayed a strong UV-fluorescence, which was considerably brighter than that of the ‘pBAD-lox-GFP’ transformed culture. Interestingly, sequence analysis (using the primer ‘GFPr2’; data not shown) revealed that ‘pBAD-lox-GFP-insN-8’ contained a trimer insertion of the 27 bp fragments.

Figure 4.5 shows a direct comparison of the UV-fluorescence intensities of the E. coli cultures that were transformed with ‘pBAD-lox-GFP’ and the various derivative constructs. It is clear from this picture that although its capacity for fluorescence is somewhat diminished from that of the parent vector ‘pBAD-GFP’, ‘pBAD-lox-GFP-insN-8’ transformed colonies fluoresce brighter than any of the other derivatives and should prove adequate for use as a visual marker of excision.

Figure 4.6 shows a western blot analysis of protein extracts from E. coli cultures (containing 0.2% arabinose to activate the PbAD promoter) transformed with ‘pBAD-GFP’ and the various loxP-containing derivatives electrophoresed on 10% SDS-PAG. The blot was probed with an anti-GFP antibody (Picture B), which identified the protein expressed from the ‘pBAD-GFP’ cassette to be approximately 27kDa in size. Although ‘pBAD-lox-GFP’ and the other mutant constructs encode proteins with extra amino acids fused to the native GFP molecule, the anti-GFP antibody also detected their expression products. The molecular weights of these proteins were clearly distinguishable on the
immunoblot, with the largest (approximately 32kDa) of these being the protein expressed from ‘pBAD-lox-GFP-insN-8’, which contains 48 additional amino acids fused to its N-terminus (compared with the parent GFP molecule expressed by ‘pBAD-GFP’). The high levels of GFP expression in \textit{E. coli} meant that each protein was clearly visible on SDS-PAGE gels stained with GELCODE® Blue Stain Reagent (shown in Picture A). This allowed relative protein concentrations to be visually compared and it is clear from this that only the constructs ‘pBAD-lox-GFP-insN-8’ and ‘pBAD-lox-GFP-ins1-6’ are capable of GFP expression levels comparable to those obtained from ‘pBAD-GFP’. The GFP protein concentrations derived from cultures containing the constructs ‘pBAD-lox-GFP’, ‘pBAD-lox-GFP-iMet’, ‘pBAD-lox-GFP-GM’ and ‘pBAD-lox-GFP-GGAG’ are considerably reduced. This result suggests that the reduced GFP fluorescence of these derivatives is most likely a consequence of translation inhibition.

\textbf{4.3.1.3. Introduction of the ‘aadA-loxP’ Cassette into ‘pBAD-lox-GFP-insN’ to make the Final Construct.}

The construction of the plasmid ‘pBAD-lox-GFP-insN’ is diagrammed in Figure 4.7. The ‘aadA-loxP’ cassette fragment was excised from pCP7-aadA-loxP by digestion with the restriction enzymes \textit{PmeI} and \textit{AvrII} and then gel-purified. The 1.4kb DNA fragment was inserted into the ‘pBAD-lox-GFP-insN-8’ plasmid digested with both \textit{PmeI} and \textit{AvrII}, causing the separation the \textit{GFP} start codon and the adjacent ‘GFP-insN’ insertion from the rest of the \textit{GFP} coding sequence in the ligated product. The correct assembly of the construct, termed ‘pBAD-lox-aadA-N-GFP’, was confirmed by restriction digest analysis (see Figure 4.10) and also by sequencing using the primers ‘BADf’, ‘Prm-Rev’ and ‘GFPr2’ (Figure 4.11).

\textbf{4.3.1.4. Use of ‘pBAD-lox-aadA-N-GFP’ as an Efficient Reporter for \textit{in vitro} Cre Activity Assays.}

Before using the construct in a plant system, ‘pBAD-lox-aadA-N-GFP’ was first tested \textit{in vitro} by exposure to purified commercial Cre recombinase. Following incubation with Cre, the plasmid was transformed into \textit{E. coli} and the resulting colonies screened for GFP-fluorescence. The result obtained is pictured in Figure 4.8, which shows a mixture of fluorescing and non-fluorescing colonies.
growing on the GFP-inducible medium. Following the optimisation of both the assay plasmid and Cre concentrations as many as 3.7% of the resulting colonies were shown to fluoresce (see Table 4.1). A non-fluorescing and a fluorescing colony were taken from such a plate and streaked on LB medium containing 50μg/ml streptomycin. Figure 4.9 shows how the non-fluorescing culture grew in the presence of the antibiotic whereas the fluorescing culture did not. This result can be explained by the occurrence of Cre-mediated recombination in the target plasmid causing the excision and loss of the loxP-flanked aadA cassette, which in turn leads to the activation of inducible GFP expression, in the case of the fluorescing culture. The non-fluorescing culture contains the unaltered target plasmid that can produce the antibiotic-detoxifying enzyme encoded by the aadA cassette via the prokaryotic-like \textit{rrn} promoter. Isolated plasmid DNA prepared from each type of culture was subjected to restriction digest analysis. Figure 4.10 demonstrates that the plasmid isolated from the fluorescing culture had an EcoRV-XbaI digest pattern similar to that of the intermediate construct ‘pBAD-lox-GFP-insN-8’, which did not contain the 1.4kb aadA cassette. Plasmid taken from the non-fluorescing culture had a digest pattern exactly like that of the parent target plasmid. Sequence analysis (using the primer ‘BADf’; Figure 4.11) provided the definitive confirmation that the plasmid isolated from the fluorescing culture (exposed to Cre) had recombined at the loxP sequence elements leading to excision and loss of the aadA cassette.

4.3.2. Development of Transgenic Tobacco Plants Expressing Cre Recombinase Targeted to the Chloroplast.

A cloned \textit{cre} gene, coding for Cre recombinase, was transformed into the nuclear genome of \textit{N. tabacum}, where its expression was controlled by the \textit{rbcS} promoter and \textit{NOS} terminator elements. Two constructs were developed: one expressed the native Cre protein in the cytoplasm (‘pRok8-Cre’ transformant); the other expressed the Cre protein fused to a transit peptide that directed the fusion protein into the chloroplast (‘pRok-Cre’ transformant).
4.3.2.1. Assembly of the *Agrobacterium* Transformation Vectors.

The cloning steps required for construction of the transformation vectors ‘pRok8-Cre’ and ‘pRok-Cre’ are detailed below and in the accompanying illustrations (Figures 4.12, 4.13 and 4.14).

4.3.2.1.1. Cloning of Cre and the *rbcS* Transit Peptide Sequences.

The cloning vector ‘pUC-Jul’ was constructed in order to facilitate assembly of the *Agrobacterium* transformation vectors. The oligonucleotides ‘Jul1’ and ‘Jul2’ were annealed to form a multiple cloning site (HindIII-XbaI-MfeI-MscI-Ncol-AatII-EcoRI-SpeI-Sacl), which was ligated into ‘pUC19’ (Yanisch-Perron *et al.*, 1985) that had been digested with EcoRI and HindIII to remove the original polylinker.

The *cre* sequence was excised and gel-purified from ‘pGST-Cre’ (unpublished plasmid), following digestion with the restriction endonucleases *Ncol* and *EcoRI*. This 1kb fragment was ligated into the *Ncol* and *EcoRI* sites (both unique to the plasmid) contained in the multiple cloning site of ‘pUC-Jul’ to produce the construct termed ‘pJul-Cre’ (used later in the construction of ‘pRok8-Cre’).

The plasmid ‘pCAR2’ (unpublished) was the source of the Rubisco SSU (*rbcS*) transit peptide sequence. The plasmid was digested using both *HindIII* and *MfeI* to release a 1.4kb fragment that was subsequently gel-purified. ‘pJul-Cre’ was also digested with *HindIII* and *MfeI*, both of which sites are located adjacent to the 5' end of the *cre* coding sequence. The transit peptide fragment was ligated into these sites to produce the final fusion construct, termed ‘pCAR-Cre’.

4.3.2.1.2. Construction of pRok-Cre.

The *Agrobacterium* binary vector ‘pRok10’ (unpublished) was digested using both *HindIII* and *Sacl*, thereby cutting a residual SSU transit peptide sequence out of the construct. The larger (>10kb vector fragment) of the two resulting DNA fragments was gel-purified. The SSU-Cre gene fusion fragment was excised from ‘pCAR-Cre’, following digestion with *HindIII* and *Sacl*. The
gel-purified binary vector fragment and the ‘pCAR-Cre’ digest were ligated, resulting in the insertion of the SSU-Cre sequence in between the rbcS promoter and the NOS terminator expression elements. E. coli transformants containing the construct ‘pRok-Cre’ were selected on LB-medium containing 50μg/ml kanamycin. Plasmid DNA minipreps were isolated from a number of clones and the correct construction was verified by restriction digest analysis (see Figure 4.15) and sequencing using the primers ‘CarSSUlf’ and ‘CreR2’ (data not shown).

4.3.2.1.3. Construction of pRok8-Cre.

The Agrobacterium binary vector ‘pRok8’ (unpublished) was digested with XbaI and SacI, with both enzymes cutting the plasmid in between the rbcS promoter and the NOS terminator expression elements. The cre gene was excised from ‘pJul-Cre’, following digestion with XbaI and SacI. Both digests were ligated and E. coli transformants containing the construct ‘pRok8-Cre’ selected on LB-medium containing 50μg/ml kanamycin. Correct plasmid construction was verified by restriction digest analysis (see Figure 4.16) and sequencing using the primer ‘CreR2’ (data not shown).

4.3.2.2. Analysis of Transgenic Tobacco Plants.

The ‘pRok-Cre’ and ‘pRok8-Cre’ binary constructs were introduced into A. tumefaciens via electroporation. Successful Agrobacterium transformation was confirmed for both strains by performing restriction digest analysis on an isolated plasmid preparation (data not shown). Agrobacterium infection was carried out separately for both strains on fresh tobacco leaf tissue. Of those infected with ‘pRok-Cre’, a total of 13 plantlets regenerated and rooted in the presence of kanamycin. Following ‘pRok8-Cre’ infection, seven kanamycin-tolerant plantlets were regenerated. All of the antibiotic-tolerant plants, labelled T1-13 (putative ‘pRok-Cre’ transformants) and C1-7 (putative ‘pRok8-Cre’ transformants), were propagated and maintained under sterile growth conditions in MS medium containing 500μg/ml carbenicillin and 100μg/ml kanamycin. Two of the plants, namely C4 and C5, did not survive the continued selection and so were not considered for subsequent analysis. The T1, T5, T12, C2 and C3 plants were also planted in soil and allowed to grow under green house conditions. Stunted growth
was observed in the case of the T12 plant but all of the other transgenic plants were morphologically identical to wild-type tobacco (data not shown). The seeds of the 5 plants were harvested and it was observed that the quantity of seeds produced by T12 was approximately ten times less than that of the others and wild-type tobacco.

All of the putative transgenic plants (T1-13, C1-3, C6 and C7) were subjected to further molecular analysis.

4.3.2.2.1. PCR Analysis.

Total DNA extracts were prepared for each of the eighteen transgenic plants (T1-13, C1-3, 6 and 7) as well as for wild-type tobacco. These were used as the templates for the PCR amplifications that facilitated the identification of transgenic ‘pRok-Cre’ and ‘pRok8-Cre’ lines as opposed to kanamycin escapes. Figure 4.17 shows the results for the PCR amplification of transgenic plant and various control-DNA preparations (94°C for 1 min; 60°C for 1 min; 72°C for 1 min; through 30 cycles) using the primers ‘CreF3’ and ‘CreR3’ that were designed to amplify an ‘internal’ cre gene fragment. A 635 bp PCR product was obtained for the plants T1-8, T10-12, C2, C3, C6 and C7, indicating the presence of the cre gene. No amplification product was obtained for T9, T13 and C1, suggesting these plants were unlikely to be transformants. Figure 4.18 shows the results of the same templates subjected to PCR amplification (94°C for 1 min; 60°C for 1 min; 72°C for 1 min; through 30 cycles) using the primers ‘CarSSU1f’ and ‘CreR3’ that were designed to amplify a sequence spanning the junction between the rbcS transit peptide sequence and the cre gene. The expected PCR product of approximately 800 bp in size was obtained for each of the plants T1, T2, T4-8 and T10-12. Unsurprisingly no such PCR product was obtained for any of the plants that contained the non chloroplast-targeted cre construct (‘pRok8-Cre’ transformants; C2, C3, C6 and C7). Interestingly these primers did not amplify a product from the T3 template, even though the ‘internal’ cre primers did produce the correctly sized PCR product. This particular transformant was therefore discarded.
4.3.2.2. Protein Analysis.

Transgenic plants were screened for expression of the Cre protein via western blot analysis. Total protein extracts were obtained from the leaf tissue of each of the transgenic plants T1, T2, T4-8, T10-12, C2, C3, C6 and C7 and also from wild-type tobacco (WT). The extracts were resolved in 8% SDS-PAGE separating gels. Having equilibrated total protein concentrations by means of visual comparison of Coomassie blue-stained protein gels (data not shown), immunoblot analysis was performed using a commercial anti-Cre antibody (Novagen; primary antibody) followed by an anti-rabbit IgG alkaline phosphatase conjugate (secondary antibody). Development of the immunoblots revealed that all of the transgenic plant extracts contained a positive signal for a 38.5kDa protein (the predicted size of Cre) that was noticeably absent for the wild-type control sample (data not shown). These immunoblots also indicated different Cre expression levels in the various transgenic lines. Figure 4.19 shows a picture of an immunoblot analysis of some of these total protein extracts and it is clear that a comparatively strong anti-Cre signal was detected for the extracts derived from the plants T1, T5 and T12. In comparison the signal obtained for the extract of T4 was much lower. Similarly a much stronger signal was obtained for the protein extract of C2 than for that of C3. Experimental repetition confirmed these findings (data not shown). The immunoblot also provided a strong signal for a positive control consisting of purified Cre protein run on the SDS-PAGE gel. The inclusion of various molecular weight markers (visualised with 0.1% Ponseau S solution) facilitated the estimation of protein size.

Chloroplasts were isolated from several transgenic lines and the resulting protein extracts subjected to western blot analysis. Protein concentrations were equalised via visual comparison on Coomassie blue-stained protein SDS-PAGE gels (data not shown). Figure 4.20 shows a picture of an immunoblot that was probed with the anti-Cre antibody and contained the chloroplast protein extracts obtained from the Cre-expressing plants T1 and C2. The blot clearly demonstrates that the chloroplast extract from T1 yielded a strong anti-Cre signal, whereas the chloroplast extract of C2 yielded a comparatively diminished signal. Interestingly, the T1-chloroplast extract signal took the form of a single-sized band, whereas the
T1-total protein extract (included on the same blot) produced a signal that suggested the presence of at least two differently sized proteins: one the same size as that indicated by the chloroplast extract signal; the other being slightly larger. No signal was obtained for the wild-type chloroplast protein extract.

4.3.2.2.3. In vitro Cre-Activity Assay of Transgenic Plant Tissue.

In vitro Cre-activity assays were performed using both total protein extracts and isolated chloroplast protein extracts. Table 4.2 outlines the results obtained from the total protein extracts taken from the leaves of the T1 and C2 transgenic plants. Two different concentrations of the Cre-assay plasmid ‘pBAD-lox-aadA-GFP’ were used in the assay, a concentration of 65ng/µl and a ten-fold lower concentration of 6.5ng/µl. Exposure to the T1 plant extract resulted in a proportion of the assay plasmid population (6.3% of the higher concentration; 11.5% of the lower concentration) undergoing site-directed recombination, an event marked by the acquisition of UV-fluorescence in the *E. coli* transformant colonies. LB-medium (containing 0.2% arabinose) plates similar to that pictured in Figure 4.8 were obtained that contained a mixture of both fluorescing and non-fluorescing colonies. Similarly to the purified Cre assay, both the fluorescing and non-fluorescing clones were tested for streptomycin resistance and both restriction digest analysis and sequence analysis was carried out on their respective isolated plasmids (illustration of results was considered unnecessary because similar results were obtained for the purified Cre assay; see Figures 4.9, 4.10 and 4.11). Fluorescing colonies were found to be streptomycin sensitive and to possess the recombined form of the assay construct. Similar results were obtained when the C2 plant extract was incubated with the higher concentration of the assay plasmid (resulting in 4% colony fluorescence).

Table 4.3 summarises the results obtained from the isolated chloroplast protein extract activity assay for the T1 and C2 transgenic plants. Two different concentrations of ‘pBAD-lox-aadA-GFP’ were used in this assay, a higher concentration of 250pg/µl and a five-fold lower concentration of 50pg/µl. Exposure of the target plasmid to the T1 chloroplast extract resulted in a proportion *E. coli* transformants (2.3% of the higher concentration; 4% of the lower
concentration) fluorescing under UV illumination. LB-medium (containing 0.2% arabinose) plates similar to that pictured in Figure 4.8 were also obtained for this assay. The colonies were analysed using the same techniques described previously for both the purified Cre and the total protein extract assays. The results (data not shown) confirmed the occurrence of fluorescing colonies to be the result of Cre-mediated recombination. Exposure of the higher plasmid concentration to the isolated chloroplast extract from C2 resulted in the production of ten times less (approximately 0.2%) fluorescing colonies compared to those produced by the T1 chloroplast extract. No fluorescing colonies were obtained when the lower plasmid concentration was used.
Figure 4.1. Introduction of a loxP Sequence Element to flank the aadA Chloroplast Expression Cassette in pCP7-aadA1. The 55 bp synthetic oligonucleotides 'loxP-aadAf' and 'loxP-aadAr' were annealed to form a double-stranded loxP fragment with SalI and AvrII 'overhangs' on either side. pCP7-aadA was digested using both SalI and AvrII and the loxP fragment was incorporated by means of a simple ligation reaction. A correctly cloned 'aadA-loxP' construct was identified using the sequence derived from the 'Prrn-Rev' primer. Note: the loxP element possesses directionality, indicated by the use of the arrow (in pink box) in the illustration.
Synthetic lox-GFP fragment was cloned into the unique NheI site of pBAD-GFP

Figure 4.2. Introduction of a loxP Sequence Element into the GFP Gene of ‘pBAD-GFP’. A gene fusion comprising of a loxP element (pink box) inserted into the GFP gene was constructed. The loxP element was translated as part of the native GFP gene, producing a fusion protein, whose UV-fluorescence capabilities were considerably diminished from those of the parent ‘pBAD-GFP’ product. A number of mutant-versions of the ‘pBAD-lox-GFP’ construct were investigated (see Figure 4.3) to correct for this deficiency and it was decided that the protein produced by ‘pBAD-lox-GFP-insN-8’ yielded sufficient fluorescence for it to be used in a visual marker system. Sequence derived from the primers indicated in the illustration showed that the construct that possessed the greatest fluorescence level contained a trimer ‘GFP-insN’ insertion.
Figure 4.3. Illustration of the various ‘pBAD-GFP’ Derivatives. These constructs were generated in order to both explain the reduction in the fluorescence intensity of GFP and to attempt to identify novel loxP insertions that might have improved fluorescence levels.
Figure 4.4. Agarose Gel Showing the PCR and Vector Intermediate Generated during the Construction of the Plasmid 'pBAD-lox-GFP-iMet'. Lane A: 1kb ladder, Lane B: 5.3kb DNA fragment, which was gel purified from the AvrII-NcoI digest of pBAD-lox-GFP, Lane C: Amplification product resulting from PCR using the primers 'GFPiMet' and 'GFP2'.
Figure 4.5. pBAD-lox-GFP Mutants Screened for UV-Fluorescence. The parent plasmid construct pBAD-GFP, pBAD-lox-GFP and its mutant constructs, plus a negative control plasmid pUC19 were transformed into *E. coli* DH5α and incubated for 48 hours at 37°C on LB agar that includes 0.2% arabinose. The cultures were photographed under a 366nm UV light source.

Figure 4.6. 10% SDS-PAGE Gel Showing Protein Extracts from *E. coli* Cultures Transformed with pBAD-GFP and the various pBAD-lox-GFP Derivatives. Picture A: Blue-stained protein gel. Picture B: Immunoblot obtained from probing with anti-GFP antibodies. Molecular weights are for the Protein Marker (Mar) that was visualised using 0.1% Ponceau S solution (Sigma®).
Figure 4.7. Assembly of ‘pBAD-lox-aadA-N-GFP’, the Target Construct for the Cre Recombinase Activity Assay. The aadA-loxP cassette was cloned into pBAD-lox-GFP-insN-8, resulting in the separation of P<sub>BAD</sub> and the GFP coding sequence. The aadA chloroplast expression cassette is in the opposite orientation to the P<sub>BAD</sub>-GFP reading frame, so aadA expression cannot be attributed to the P<sub>BAD</sub> promoter. The aadA expression cassette is flanked on either side by a loxP recombination element (pink boxes). Cre recombinase activity results in the removal of the aadA cassette and the rejoicing of P<sub>BAD</sub> with the correct GFP reading frame. Correct construct assembly was confirmed by using the primers indicated in the illustration to obtain the sequence across the cloning junctions.
Figure 4.8. Chloroplast-Localised Cre Activity Assay. The assay construct pBAD-lox-aadA-N-GFP was exposed in vitro to purified Cre activity. The plasmid was subsequently transformed into E. coli DH5α and incubated for 48 hours at 37°C on LB agar that includes 0.2% arabinose. The colonies were photographed under a 366nm UV light source. A number of colonies fluoresced bright green, indicating Cre Recombinase activity, while others did not. Similar pictures were obtained (not shown) for assays involving chloroplast protein extracts isolated from the leaf tissue of the transgenic tobacco clone 'T1' as well as for total plant protein extracts.
1. pUC19 (negative control)  2. pBAD-lox-GFP-insN-8  3. pBAD-lox-aadA-N-GFP
4. ‘Non-fluorescing’ Clone  5. ‘Fluorescing’ Clone

**Figure 4.9.** The five cultures numbered above, including both a fluorescing and non-fluorescing colony from the plate pictured in Figure 4.8, were streaked on both LB medium containing 50µg/ml streptomycin and LB medium containing 0.2% arabinose. **Picture A** is of the streptomycin-selection plate following an overnight incubation at 37°C. **Picture B** is the arabinose-containing plate illuminated with 366nm UV light following a 48-hour incubation at 37°C.

**Figure 4.10.** Restriction Analysis of pBAD-lox-aadA-N-GFP following Exposure to Cre Recombinase Activity. **Lane1:** ‘Unexposed’ pBAD-lox-aadA-N-GFP digested with both EcoRV and XbaI. **Lane2:** ‘Unexposed’ pBAD-lox-GFP-insN-8 digested with both EcoRV and XbaI. **Lane3:** Target plasmid, isolated from a ‘non-fluorescing’ clone subjected to Cre activity (see Figures 4.8 and 4.9), digested with both EcoRV and XbaI. **Lane4:** Target plasmid, isolated from a ‘fluorescing’ clone subjected to Cre activity, digested with both EcoRV and XbaI.
Figure 4.11. Sequence Confirmation of Cre-mediated Recombination in the Plasmid ‘pBAD-lox-aadA-N-GFP’. The blue-colored nucleotide bases represent the 34 bp recombination element loxP. The directionality of this element is determined by the orientation of the spacer region (underlined bases). 1: The sequence of ‘pBAD-lox-aadA-N-GFP’ that was obtained using the primer ‘BADf’. The \textit{psbA} 3’ terminator element of the \textit{aadA} cassette was located downstream of one \textit{loxP} site; 2: The sequence of ‘pBAD-lox-aadA-N-GFP’ that was obtained using the primer ‘GFPr2’. The promoter element of the \textit{aadA} cassette was located upstream of the other \textit{loxP} site; 3: The sequence of the plasmid isolated from a UV-fluorescing colony (described in Figures 4.8 and 4.9) confirms site-specific recombination at the \textit{loxP} elements with the excision of the intervening \textit{aadA} cassette.
Figure 4.12. Flowchart for the Construction of pCAR-Cre. The cre gene was cloned as an EcoR1–NcoI restriction fragment into the polylinker of the cloning vector pUC-Jul. The rbcS transit peptide sequence (SSU-TP) was subsequently cloned into this vector to the 5' of cre, having been purified as a HindIII–MfeI restriction fragment from the pCAR2 construct. The final construct contains the cre gene fused to the SSU-TP with restriction sites suitable for subsequent cloning steps.
Figure 4.13. Construction of pRok8-Cre, the *Agrobacterium* Vector Construct used to introduce the Cre Expression Cassette (Minus Chloroplast Transit Peptide) into the Nuclear Genome of *N. tabacum*. Cre expression was driven by the rbcS promoter and regulated by the NOS terminator. The primers used to confirm the construct sequence and to screen transformed plants by PCR are shown in blue.

Figure 4.14. Construction of pRok-Cre, the *Agrobacterium* Vector Construct used to introduce the Cre Expression Cassette (Includes Transit Peptide) into the Nuclear Genome of *N. tabacum*. The rbcS promoter and the NOS terminator elements controlled SSU-cre expression. The primers used to confirm the sequence of the construct and to screen transformed plants by PCR are included.
Figure 4.15. Restriction Digest Analysis of the pRok-Cre Construct. The parent and final plasmid constructs were digested using both HindIII and SacI. Lane 1: pRok10 digest; Lane 2: pCAR-Cre digest; Lane 3: pRok-Cre.

Figure 4.16. Restriction Digest Analysis of the pRok8-Cre Construct. The parent and final plasmid constructs were digested using both XbaI and SacI. Lane 1: pRok8 digest; Lane 2: pRok8-Cre.
Figure 4.17. PCR Analysis of both pRok-Cre and pRok8-Cre Transformed Tobacco Plants using Two ‘Internal’ cre Primers. Total DNA extracts were prepared for all putative transgenic plants and used as the templates for PCR. Both the ‘CreF3’ and ‘CreR3’ primers (see Figures 4.13 and 4.14) were chosen to hybridise to DNA sequence contained within the cre coding region. As a control a wild-type DNA preparation (WT) was subjected to the PCR test, as were the purified plasmid constructs pUC19 (pX; negative control), pRok8-Cre (pY) and pRok-Cre (pZ).

Figure 4.18. PCR Analysis of both pRok-Cre and pRok8-Cre Transformed Tobacco Plants using One ‘Internal’ cre Primer and another positioned in rbcS Transit Peptide Sequence. The primers ‘CarSSUlf’ and ‘CreR3’ (see Figure 4.14) only yielded a PCR product (approximately 800bp) where the transgenic plant contained a copy of the cre gene fused to the transit peptide. This molecular profile distinguished plants transformed with pRok-Cre from those transformed with pRok8-Cre.
Table 4.1. *In vitro* Activity of Purified Cre Recombinase on pBAD-lox-aadA-N-GFP. Following exposure to purified Cre the target plasmid was transformed into *E. coli* DH5α and incubated for 48 hours at 37°C on LB agar that contains 0.2% arabinose. All colonies were illuminated using 366nm UV light and screened for fluorescence that would indicate the occurrence of site-directed recombination on the plasmid. The inert pUC19 plasmid was also exposed to Cre and transformed to act as a control.

<table>
<thead>
<tr>
<th>Cre Concentration</th>
<th>Total Number of Colonies</th>
<th>Number that Fluoresced</th>
<th>% Fluorescing</th>
</tr>
</thead>
<tbody>
<tr>
<td>75mU/μl Cre; 6.5ng/μl pBAD-lox-aadA-N-GFP</td>
<td>~1000</td>
<td>37</td>
<td>~3.7%</td>
</tr>
<tr>
<td>0mU/μl Cre; 6.5ng/μl pBAD-lox-aadA-N-GFP</td>
<td>~1000</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>75mU/μl Cre; 6.5ng/μl pUC19</td>
<td>~1000</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>0mU/μl Cre; 6.5ng/μl pUC19</td>
<td>~1000</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 4.2. *In vitro* Cre-Activity of Transgenic Plant Total Protein Extracts. Two different concentrations of the Cre-target plasmid pBAD-lox-aadA-GFP were exposed to total protein extracts from the leaves of transgenic plants ‘T1’ and ‘C2’ and wild-type (WT) tobacco. Cre recombinase activity was implied by the presence of UV-fluorescent colonies following transformation into *E. coli* DH5α.

<table>
<thead>
<tr>
<th>Extract Type</th>
<th>Total Number of Colonies</th>
<th>Number that Fluoresced</th>
<th>% Fluorescing</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘T1’ Total Extract; 65ng/μl pBAD-lox-aadA-N-GFP</td>
<td>238</td>
<td>15</td>
<td>6.3%</td>
</tr>
<tr>
<td>‘T1’ Total Extract; 6.5ng/μl pBAD-lox-aadA-N-GFP</td>
<td>52</td>
<td>7</td>
<td>11.5%</td>
</tr>
<tr>
<td>‘C2’ Total Extract; 65ng/μl pBAD-lox-aadA-N-GFP</td>
<td>276</td>
<td>11</td>
<td>4%</td>
</tr>
<tr>
<td>‘C2’ Total Extract; 6.5ng/μl pBAD-lox-aadA-N-GFP</td>
<td>22</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>‘WT’ Total Extract; 65ng/μl pBAD-lox-aadA-N-GFP</td>
<td>~500</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>‘WT’ Total Extract; 6.5ng/μl pBAD-lox-aadA-N-GFP</td>
<td>18</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Total Number of Colonies</td>
<td>Number that Fluoresce</td>
<td>% Fluorescing</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>'T1' Chloroplast Extract; 250pg/µl pBAD-lox-aadA-N-GFP</td>
<td>~1000</td>
<td>23</td>
<td>~2.3%</td>
</tr>
<tr>
<td>'T1' Chloroplast Extract; 50pg/µl pBAD-lox-aadA-N-GFP</td>
<td>177</td>
<td>7</td>
<td>4%</td>
</tr>
<tr>
<td>'C2' Chloroplast Extract; 250pg/µl pBAD-lox-aadA-N-GFP</td>
<td>~1000</td>
<td>2</td>
<td>~0.2%</td>
</tr>
<tr>
<td>'C2' Chloroplast Extract; 50pg/µl pBAD-lox-aadA-N-GFP</td>
<td>122</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>'WT' Chloroplast Extract; 250pg/µl pBAD-lox-aadA-N-GFP</td>
<td>~1000</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>'WT' Chloroplast Extract; 50pg/µl pBAD-lox-aadA-N-GFP</td>
<td>317</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 4.3. *In vitro* Cre-Activity of Transgenic Plant Chloroplast Protein Extracts. Two different concentrations of the Cre-target plasmid pBAD-lox-aadA-GFP were exposed to chloroplast protein extracts from the leaves of transgenic plants 'T1' and 'C2' and wild-type (WT) tobacco. Cre recombinase activity was implied by the presence of UV-fluorescent colonies following transformation into *E. coli* DH5α.
4.4. Discussion.

The overall aim of this chapter was to develop an efficient system for the removal of marker genes from the plastid genome of higher plants. In light of the recent advances made in the application of site-specific recombination systems in heterologous organisms, including plants, the Cre/loxP system was chosen to perform such a task. Transgenic plants were developed that expressed the Cre protein from the nuclear genome. Cre was targeted to the chloroplast by fusing it with an N-terminal plastid-targeting signal (the SSU transit peptide). The construct ‘pBAD-lox-aadA-N-GFP’ was developed as an assay plasmid for Cre recombinase activity and ultimately for use as part of a plastid transformation vector. Cre/loxP-mediated recombination of the construct resulted in the excision of the aadA marker gene with simultaneous activation of GFP expression.


The construct ‘pBAD-lox-aadA-N-GFP’ contains two protein-coding sequences, those of aadA and GFP. Expression of the marker gene aadA was controlled by the chloroplast Prrn promoter and psbA 3’ elements, with the entire cassette flanked on either side by a loxP recombination site. The insertion of this loxP-flanked cassette into the GFP reading frame eliminated the possibility of GFP expression from the P_{BAD} promoter. Cre-mediated recombination at the loxP sites resulted in the site-specific excision of the aadA cassette. Such an event restored the GFP reading frame with only a single loxP element persisting.

It was a necessary inconvenience of the system that the loxP element was translationally fused to the native GFP coding sequence. Unfortunately, this fusion led to a dramatic reduction in UV-fluorescence capacity, as observed in the ‘pBAD-lox-GFP’-transformed E. coli cultures. The development of a range of derivative constructs led to the eventual conclusion that the reduced fluorescence was a consequence of decreased GFP protein expression. This was most probably due to the formation of an mRNA stem-loop structure because of the presence of
inverted-repeat sequences within the *loxP* element, which may have interfered with ribosome binding and consequently translational efficiency (Liu *et al.*, 1998). In the case of both the ‘pBAD-lox-GFP-ins1-6’ and ‘pBAD-lox-GFP-insN-8’ constructs, the introduction of specific short sequences adjacent to the *loxP* site resulted in increased GFP expression. This conclusion was substantiated by the very large difference in GFP expression levels that was clearly visible in the blue-stained protein gels (see Figure 4.6A) and also by the optical comparison of the arabinose-induced UV-fluorescence produced by *E. coli* cultures (see Figure 4.5). The additional introduced sequences may have caused relaxation of the stem-loop structure and thus facilitated more efficient ribosome binding and translation.

**4.4.2. Transgenic Tobacco Plants Expressing Cre.**

Transgenic tobacco plants were developed by *Agrobacterium*-mediated transformation of leaf tissue using two different binary constructs. The introduction of ‘pRok8-Cre’ or ‘pRok-Cre’ led to transformants successfully expressing Cre recombinase from the nuclear genome, with the latter construct producing a protein that was targeted to the chloroplast via a cotranslated transit peptide. Of the 13 ‘pRok-Cre’-infected plants regenerated under kanamycin selection, a total of ten were shown to have incorporated the construct, by using PCR analysis (see Figures 4.17 and 4.18). Similarly, four out of the seven ‘pRok8-Cre’-infected regenerants contained the nuclear-integrated construct. The remainder of the plants were considered to be escapes of the kanamycin selection system.

Western blot analysis of the transgenic plants (see Figure 4.19) showed the fact that different Cre expression levels existed between plant lines containing the same construct. Such an occurrence is a common feature of the *Agrobacterium*-mediated transformation system due to the lack of positional control for transgene insertion into the nuclear genome. This may result in epigenetic gene silencing effects, which may vary depending on the genomic location of the inserted transgene.
Observation of transgenic plant morphology revealed one of the plantlets (T12) to possess stunted growth and reduced seed production, which were particularly obvious. These results may be due to the interruption of a native gene coding sequence or expression control element. Alternatively, they could be a consequence of transgene hyperexpression interfering with normal plant physiology. Time constraints prevented us from investigating this plant line any further. All of the other Cre expressing plants (both for 'pRok8-Cre' and 'pRok-Cre') appeared to have a growth and reproductive phenotype similar to that of wild-type tobacco.

These results can also be used to highlight one of the advantages of using chloroplast transformation techniques in transgene expression systems. Because transgene introduction occurs via homologous recombination in the chloroplast, the exact insertion site is controlled by the choice of chloroplast flanking sequence. Therefore, possible native gene-knockout and epigenetic effect considerations can be largely ignored.

The successful targeting of the nuclear expressed Cre protein into the chloroplast was strongly supported by two lines of evidence. Firstly, western blot analysis (see Figure 4.20) demonstrated that isolated chloroplast protein extracts from the ‘T1’ clone, which had an integrated SSU-cre construct (‘pRok-Cre’ transformant), contained high concentrations of the Cre protein. Neither the ‘C2’ clone, which possessed the non-targeting cre construct (‘pRok8-Cre’ transformant), nor the wild-type chloroplast protein extracts contained levels of the protein that were identifiable by the anti-Cre antibody. A detailed examination of the ‘T1’ total protein extract revealed the presence of a higher molecular weight band in addition to a band common to both the ‘C2’ total protein and the ‘T1’ chloroplast protein extracts. It is possible that this higher molecular weight signal represents the cytosolic localised Cre fusion protein with its unprocessed transit peptide domain. The transit peptide domain is predicted to be cleaved from the Cre domain, once it enters the chloroplast (Keegstra and Cline, 1999; Pinnaduwage and Bruce, 1996). The second line of evidence demonstrating Cre localization within chloroplasts comes from the results of the in vitro Cre activity assays using protein extracts
from transgenic plant tissues (see Tables 4.2 and 4.3). Although their total protein extract activities (identifiable by ‘pBAD-lox-aaaA-N-GFP’ recombination) were shown to be similar, the isolated chloroplast protein extract activity of ‘T1’ was approximately ten times greater than that of ‘C2’. The minimal (0.2%) activity detected in the ‘C2’ chloroplast protein extracts is probably due to residual cytosolic protein contaminating in the chloroplast preparation.

4.4.3. In vitro Cre Activity Assays.

In vitro recombinase activity assays were performed separately using either purified commercial Cre enzyme, transgenic plant total protein extracts or isolated chloroplast protein extracts. Two different target-plasmid concentrations were used in the assays and the results showed that recombination frequencies were influenced by plasmid concentration. In general the lower plasmid concentration displayed a higher recombination frequency when a constant amount of Cre-containing protein extract was used. Both the upper and lower ranges of plasmid concentrations that could be used in each assay were dependent on the efficiency of the *E. coli* transformation technique. An optimal number (~100 to ~1000) of colonies (that were clearly separated) was required. The *E. coli* transformation efficiencies were influenced by compounds present in the plant tissue extract solutions, which meant that the optimal plasmid concentration range varied for each assay type (e.g. *E. coli* transformation was approximately 20 times less efficient for total plant protein extracts than for the commercially-purified Cre (see Tables 4.1 and 4.2)). Therefore, it was not possible to provide a direct comparison of Cre recombinase activity from the different sources by this system alone. It is possible that Cre-specific activity may be altered in plant systems, but the results do suggest that activities were sufficiently high to allow in vivo (in planta) recombination to occur.

4.4.4. Cre/loxP-mediated Excision in Chloroplasts.

Unfortunately, while the work described in this Chapter was in progress, two papers were published which demonstrated the feasibility of Cre/loxP-mediated excision of marker genes in chloroplasts. Hajdukiewicz *et al.* (2001) demonstrated antibiotic resistance marker gene removal from the plastome of
tobacco and simultaneous GFP reporter gene activation. A similar experiment is described in this chapter. Cre/loxP mediated excision was shown to occur early in plant development and homoplasmy is reached quickly in the absence of selective pressure (Hajdukiewicz et al., 2001). Selectable marker-free plants, with the nuclear-expressed Cre gene removed, were obtained as a subpopulation of the first seed generation. Alternative excision events were also observed that were most likely a consequence of the unique features of the plastid's own recombination machinery. The presence of Cre recombinase in the chloroplast was apparently capable of stimulating homologous recombination between the transgene promoter element (Frrn) and the cognate plastome Prrn sequence. Also, illegitimate recombination between loxP and a plastome “hotspot” resulted in the deletion of approximately 500 bp of plastid sequence containing a tRNA gene. Fortunately transplastomic plants, homoplasmic for the correct excision event, were stable and no additional recombination events occurred.

In the second report Corneille et al. (2001) demonstrated Cre-mediated excision of a codA-aadA gene cassette flanked by loxP sites. Cre recombinase was targeted to the chloroplast by translational fusion with the Rubisco small subunit (SSU) transit peptide and introduced either into somatic cells by Agrobacterium-mediated transformation or germline cells via pollen. Although both systems were very efficient for loxP-flanked DNA removal, the somatic excision method was regularly accompanied by a large deletion, absent from those lines derived following germline excision. This suggests that Cre recombinase may interact with the plastid recombination system in a tissue-specific manner.

4.4.5. Future Work.

The results described in this chapter will be used towards the completion of an efficient chloroplast marker gene removal system. A number of steps remain to be implemented. The ‘pBAD-lox-aadA-N-GFP’ cassette must be modified for chloroplast expression and transformed into the plastid genome. Transformants can be selected on the basis of acquired resistance to both spectinomycin and streptomycin. Chloroplast transformation is now routinely achieved for N. tabacum (see General Introduction). Once homoplasmy has been confirmed, this
plant will then be crossed with the \textit{N. tabacum} strain expressing Cre recombinase in the chloroplast ('T1').

The results obtained by both Hajdukiewicz \textit{et al.} (2001) and Corneille \textit{et al.} (2001), who both introduced a \textit{loxP-aadA} cassette into the inverted repeat region of the \textit{N. tabacum} chloroplast genome, suggest that delivery of Cre recombinase via pollination was the most efficient route for achieving marker gene excision. In these studies it was observed that when the transplastomic strain (including the marker gene cassette) was directly transformed with the fused Cre-TP cassette, the recombined (excised marker gene) plastid genome only reached homoplasmy after several rounds of regeneration. A sexual-cross resulted in marker-free transplastomic plants by the first seed generation, indicating that the chloroplast-localised Cre enzyme was active early in plant development. These publications also highlighted the frequent occurrence of Cre-stimulated general recombination between transgene expression elements and cognate sequences in the chloroplast genome. Therefore, for the present project, it is advisable that such findings are taken into account when deciding on the exact plastome location for the \textit{loxP} marker gene cassette.
GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

The primary aim of the investigations described in Chapter 2 was to facilitate the construction of *Brassica*-specific plastid transformation vectors. The restriction fragment clone banks developed for each of the enzymes HindIII, EcoR1, SalI and XbaI, gave a total coverage of 97% of the *B. napus* plastid genome with significant overlap. Extensive sequencing of a number of these cloned DNA fragments lead to the accumulation of a total of 56.6kb of chloroplast DNA sequence representing almost half of the unique sequence of the genome. In addition to facilitating transformation vector design, the sequence data also allowed detailed phylogenetic analyses to be performed in conjunction with a number of other higher plant plastid DNA sequences, available in the public databases. The analyses were performed on a number of levels which included analyses based on individual plastid gene coding sequences, concatenated coding sequences and non-coding intergenic sequences. Phylogenetic trees were constructed and provided some interesting results about the diversification of the angiosperm lineage. The results confirmed the close ancestry that exists between *B. napus* and *A. thaliana* and the occurrence of an explosive species radiation approximately 100 million years ago. Phylogenetic tree calibration (taking the monocot/dicot split to be at 200 (± 40) million years) provided an *Arabidopsis-Brassica* divergence date of 14.46-21.69 million years, an estimate that is in approximate agreement with the conclusions of Yang *et al.* (1999a). The high levels of sequence similarity observed between *B. napus* and *A. thaliana* suggests that *Brassica* biotechnology (including plastid genome transformation) will likely benefit from the availability of the *Arabidopsis* nuclear genome sequence. The considerable mass of *B. napus* plastid genome sequence will also be useful for the study of the numerous other *Brassica* crop species.

Since the Martin *et al.* (1998) publication, the entire plastid genome sequence of several other higher plant species have become available, including those from *A. thaliana* (Sato *et al.*, 1999), *O. elata* (Hupfer *et al.*, 2000), *L. japonicus* (Kato *et al.*, 2000), *T. aestivum* (Ogihara *et al.*, 2002) and *S. oleracea* (Schmitz-Linneweber *et al.*, 2001). Now that a diverse range of higher plant
plastid genome sequences is available, it will be possible to undertake a multi-gene study, similar to that described in Martin et al. (1998). Such a study should prove valuable in answering a number of questions surrounding the recent evolution of the plastid genome.

Two transformation vectors were constructed to facilitate the targeted insertion of foreign DNA sequences into different locations of the *B. napus* plastid genome. The vector pZB1 facilitates the insertion of an *aadA* expression cassette into the *rbcl-aacD* intergenic region by homologous recombination. In contrast, pPC1 was designed to facilitate foreign gene insertion into the *trnV-rps12 (3')* intergenic sequence by means of a "binding type" antibiotic resistance selection system. Because an efficient regeneration system for *B. napus* chloroplast transformation has yet to be developed, the Brassica-specific pPC1 construct was tested in a tobacco system. In spite of the heterogeneity existing between the vector and target sequences, two putative plastid transformants were recovered. The point mutation engineered into the *16S rRNA* gene in pPC1 in order to confer spectinomycin resistance was shown to have been successfully integrated into the tobacco plastid genome. Sequence analysis of the integrated region showed that the tobacco plastid genome also contained the two Brassica-specific single nucleotide substitutions located on either side of the antibiotic resistance-conferring point mutation. PCR analysis also revealed that neither transformant contained the pPC1 intergenic multiple cloning site or the streptomycin point mutation, located within the *rps12 (3')* coding sequence. This suggests that homologous recombination was most likely confined to within the *16S rRNA* coding region, a gene that is highly conserved (99.5%; see Table 2.5) between the genomes of *B. napus* and *N. tabacum*. The *rps12 (3')* gene was found to be less conserved (98.4%; see Table 2.5) between the two species and this may have reduced the probability of a homologous recombination event occurring in this region. However, if statistically reliable conclusions are to be made about homologous recombination frequencies, many more transformants will need to be generated using both spectinomycin and streptomycin selection systems.
The true value of both pPC1 and pZB1 will be realised once an efficient regeneration system for *Brassica* chloroplast transformation has been developed. It is expected that each will function efficiently as a species-specific plastid transformation vector. However, provided sufficient homology exists between the vector flanking sequences and the target genome, they may also be useful for plastid transformation in related species at least within the *Brassica* lineage (e.g. *Brassica oleracea*) and possibly in other species within the Brassicaceae. In this context however, it is likely that pZB1 will be less useful as a "universal vector" than pPC1, because both the *rbcL* and *accD* coding and intergenic sequences diverge greatly between species, even those as closely related as *B. napus* and *A. thaliana* (Tables 2.5 and Figure 2.17).

Chapter 4 described the development of a site-specific recombination system for removal of marker genes in transformed plastids. The Cre recombinase expression construct, which included a transit peptide sequence that targets the protein into the chloroplast, was inserted into the nuclear genome of tobacco by *Agrobacterium*-mediated transformation. Western blot analysis and an *in vitro* Cre activity assay indicated that the resulting transgenic plants expressed a functional Cre recombinase that was localized in chloroplasts. The *in vitro* assay was based on Cre-mediated excision of an *aadA* cassette from the plasmid construct “pBAD-lox-aadA-N-GFP” which leads to activation of GFP expression. This construct will ultimately be incorporated as part of the selection cassette in various plastid transformation vectors e.g. pZB1. Excision of the selection cassette can then be effected in homoplasmic plastid transformants either by a sexual cross with a transgenic line carrying a nuclear Cre recombinase transgene or by direct *Agrobacterium*-mediated transformation of the homoplasmic line with the Cre transgene. Moreover because excision leads to GFP expression, its occurrence is easy to verify. A major advantage of this approach to marker gene removal is that excision is highly controllable and the system would be expected to function in a species-independent manner. This technology is therefore likely to be widely applied for removal of antibiotic resistance genes in transplastomic crop plants.
During the assembly of the “pBAD-lox-aadA-N-GFP” assay construct, it was observed that the insertion of a loxP recombination element into the 5’ end of the GFP open reading frame, resulted in the dramatic reduction of GFP activity. This result was surprising because the visual marker protein had previously been shown to be extremely compatible with larger fusion domains (Kolb and Siddell, 1996). The subsequent investigations, consisting of the generation of a number of mutant constructs, revealed that the reduction in activity was due to the inhibition of mRNA transcription. This was most probably a consequence of a stem-loop structure forming between the 13 bp inverted repeats in the loxP element, inhibiting the assembly of the transcription initiation complex. One particular mutant, “pBAD-lox-GFP-insN-8”, was demonstrated to express GFP at sufficiently high levels for it to yield a clearly distinguishable phenotype in E. coli. With the Cre/loxP recombination system and the heterologous expression of GFP both being such important tools in modern molecular research, the development of an active GFP construct that is capable of Cre-mediated recombination, is certain to be applicable to many disciplines.


Murai, R., Murai R., Murai


Tiler, K. and Link, G. (1993a). Sigma-like transcription factors from mustard (Sinapis alba L.) etioplast are similar in size to, but functionally distinct from, their chloroplast counterparts. Plant Mol. Biol. 21, 503-513.


the Hamamelidaceae: Late Cretaceous (Turonian) inflorescences and fruits of