LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

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.

Mitochondrial Genes Associated with Cytoplasmic Male Sterility in *Lolium perenne*

A thesis submitted for the degree of Doctor of Philosophy at the University of Dublin, Trinity College.

By

Paul McDermott

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

March, 2008.

TRINITY COLLEGE
2 7 MAY 2008
LIBRARY DUBLIN
THESIS
8378

Declaration

I hereby certify that this thesis, submitted to the University of Dublin for the examination for the degree of Doctor of Philosophy, has not been submitted as an exercise for a degree at any other university. I also certify that this thesis is entirely my own work.

This thesis shall be made available for consulation within the University library and can either be photocopied or loaned to other libraries for the purpose of consultation.



Paul Mc Dermott

March, 2008.

To Anne, Ciara and Fionn

and

In memory of my late Father Brendan and my late Uncle Patsy

"Where I am, I don't know,
I'll never know, in the silence you don't know,
you must go on,
I can't go on,
I'll go on": **Samuel Beckett - The Unnamable (1953)**

Acknowledgements

I would firstly like to thank my supervisor, Prof. Tony Kavanagh, for giving me the opportunity to conduct my research thesis in Plant Molecular Genetics, for his encouragement and support throughout the project, and for his much appreciated help in the editing of this thesis.

Thanks also to Prof. David McConnell, who initially sowed the seed to undertake a line of research in the Genetics Department, and I'm very grateful to him for his initial advice and enthusiasm.

It would not have been possible to undertake a research thesis without the backing of the Chief Technician – Dave Sullivan. I'll always be deeply indebted to Dave in allowing me the opportunity to utilise departmental time on my thesis. I would also like to thank Brenda who covered for me on numerous occasions, particularly during the final stages of the project when time was at a minimum – thanks a million to you both.

A special word of thanks to Alan for his companionship, his sense of humour and above all for being a good friend over the last number of years.

I would also like to thank all of the gang in the Plant Molecular Genetics laboratory, both past and present, especially Alan, Jim, Nga, Sinead and Seamus.

I wish to thank my wife Anne who has shown great support, endurance and tolerance for my inactivities in doing the household duties, when instead, I was either slogging it out in the laboratory, or at home writing my thesis. I would like to pay a special tribute to my mother Margaret for being my 'Rock of Gibraltor,' who has shown great encouragement throughout my educational life. Finally, I would like to thank my two sisters Miriam and Margaret for their help and moral support, particularly during those countless times when the going got tough.

I dedicate this thesis to my wife Anne and to my two children Ciara and Fionn.

Table of Contents

Page

Summary i

Abbreviations iii

CHAPTER 1

General Introduction

1.1 The Mitochondrial Genome in Plants 1

1.1.1 Gene content of plant mitochondrial genomes – an overview 2

1.1.2 Mitochondrial genes of non-mitochondrial origin 3

1.1.3 Replication of plant mitochondrial DNA 4

1.1.4 Recombination and the multipartite structure of plant mitochondrial genomes 5

1.1.5 Recombination and mitochondrial mutations 6

1.1.6 Mitochondrial plasmids 7

1.1.7 Gene expression in plant mitochondria 8

1.1.7.1 RNA editing in plant mitochondria 9

1.1.7.2 The mechanism of RNA editing site selection 10

1.2 Cytoplasmic Male Sterility (CMS) in flowering plants 11

1.2.1 Mitochondrial genes associated with cytoplasmic male sterility 12

1.2.2 Chimeric genes and CMS 13

1.2.2.1 How does CMS affect male fertility but, in general, not vegetative growth? 14

1.2.3 CMS and programmed cell death 16

1.3 Perennial Ryegrass 17

1.3.1 The Ryegrass-Fescue *(Lolium-Festuca)* complex 18

1.3.2 CMS in perennial ryegrass 18

1.4 Thesis Aims 20

CHAPTER 2

General Materials and Methods

2.1 Materials 21

2.1.1 General chemicals 21

2.1.2 Plant cell culture reagents 21

2.1.3 Molecular biology 22

2.1.3.1 Cloning reagents 22

2.1.3.2 Radiolabelling 22

2.1.3.3 PCR amplification (Taq) 23

2.1.3.4 Reverse transcription (RT)-PCR 23

2.1.3.5 Protein expression analysis 23

2.1.3.6 Stock solutions 24

2.2 Molecular Biology Methods 26

2.2.1 Electrophoresis methods 26

2.2.2 Transformation of *E.coli* Competenet Cells 26

2.2.3 Plasmid DNA isolation for cloning and sequencing 27

2.2.4 Sequencing 28

2.2.5 Southern blotting 29

2.2.6 Northern blotting 30

2.2.7 Reverse transcription (RT)-PCR 31

2.2.8 Total cellular DNA extraction 31

2.2.9 Total RNA extraction 32

2.2.10 Protein analysis 33

2.2.10.1 Electrophoresis 33

2.2.10.2 Protein expression in *E.coli* 33

2.2.10.3 Purification of an insoluble protein by electro-elution 33

2.2.10.4 Generating polyclonal antibodies in rabbit 34

2.2.10.5 Western blotting 34

2.2.10.6 Affinity purification of antibodies 35

CHAPTER 3

Analysis of a Novel CMS Gene, *orfC9*, in the Mitochondrial Genome of *Lolium perenne*

3.1 Introduction 37

3.1.1 CMS chimeric genes associated with ATP synthase loci 37

3.1.2 Cytochrome oxidase genes associated with CMS chimeric orfs 39

3.1.3 Restoration of fertility 40

3.2 Chapter Aims 41

3.3 Materials and Methods 43

3.3.1 Materials 43

3.3.1.1 *Lolium perenne* lines 43

3.3.1.2 Primers 43

3.3.2 Methods 44

3.2.2.1 PCR amplification of *atp9*, *orfC9* and *orf-L* fragments 44

3.3.2.2 Recombinant protein expression 44

3.3.2.3 Sequencing 44

3.3.2.4 Extraction of total protein from leaf tissue 44

3.3.2.5 Extraction of enriched mitochondrial protein 45

3.3.2.6 Generation of polyclonal antibodies recognising an ORFC9 fusion protein 45

3.3 Results 46

3.4.1 Analysis of *atp9*-associated transcripts in fertile, sterile and restored *L. perenne* lines 46

3.4.2 Investigation of *atp9* and *orfC9* editing in sterile and restored *L. perenne* lines 46

3.4.3 Determination of *atp9*, *orfC9* and *orf-L* copy number in CMS and fertile mtDNA 47

3.4.4 Expression of recombinant ORFC9 protein in *E. coli.* 48

3.4.5 Western blot detection of ORFC9 50

3.5 Discussion 51

CHAPTER 4

Identification of an integrated linear plasmid-like element in the mitochondrial genome of a CMS line of *L. perenne*

4.1 Introduction 55

4.1.1 Linear mitochondrial plasmids in plant and fungi 55

4.1.2 Replication of linear plasmids 56

4.1.3 Integration of linear plasmids into the mitochondrial genome 57

4.1.4 Phenotypic consequences of linear plasmid integration 57

4.1.5 Transmission of linear mitochondrial plasmids 58

4.2 Chapter Aims 59

4.3 Materials and Methods 60

4.3.1 Plant material 60

4.3.2 Molecular techniques 60

4.3.3 Sequencing of clones from CMS-St and the fertile line LPS21 60

4.3.4 Primers used in sequencing and analysis of *L. perenne* mtDNA 60

4.3.4.1 Oligonucleotide primers used in the sequencing of pCMS subclones 60

4.3.4.2 Oligonucleotide primers used in the sequencing of pLMF12 subclone 63

4.3.4.3 Primers used in the amplification of Dpo-like, Rpo-like and the intergenic sequences from fertile lines of *L. perenne* 63

4.3.4.4 Primers used to amplify fragments from a cDNA template in sterile (CMS9B290) and restored (Rst1) lines 66

4.3.4.5 Primers used for generating protein expression constructs 67

4.3.5 Nucleotide sequencing and sequence analysis 67

4.3.6 RNA analysis 68

4.3.7 Protein expression 68

4.3.8 PCR conditions 68

4.3.8.1 PCR amplication using total cellular DNA template 68

4.3.8.2 PCR amplication of fragment from orf1 for protein expression 68
4.3.8.3 PCR amplification from the cDNA template 68
4.3.8.4 Long range PCR amplification from total DNA template 69
4.3.8.5 Proteinase K treatment of total cellular DNA 69
4.4 Results 70
4.4.1 Construction of subclones spanning LpCMSi and the homologous insertion site locus in the fertile *L. perenne* line LPS21 70
4.4.2 Identification of open reading frames in LpCMSi 72
4.4.2.1 ORF1 encodes a family B-type DNA polymerase 72
4.4.2.2 ORF2 encodes a single subunit, phage-like RNA polymerase 74
4.4.2.3 ORF3 is unique 74
4.4.3 PCR detection of CMSi-like sequences in fertile lines of *L. perenne* 75
4.4.4 Evidence that the LpCMSi-like sequences in fertile lines of *L. perenne* may exist as extrachromosomal replicons 76
4.4.4.1 Verification of endophyte infection in the *L. perenne* line FNz 77
4.4.5 Sequences homologous to ORF1 and ORF2 in plant mitochondrial genomes 78
4.4.6 Phylogenetic analysis of Dpo- and Rpo-like sequences in plants and fungi 79
4.4.7 Dpo and Rpo transcripts are detected in *L. perenne* lines containing the CMS cytoplasm 80
4.4.8 Expression of an LpDpo fusion protein in *E. coli* 81
4.4.8.1 LpDpo fusion protein expression using pTwin-Impact 81
4.4.8.2 LpDpo fusion protein expression using pET32a 82
4.5 Discussion 84

CHAPTER 5

Expression of a CMS associated *orfC9* sequence in Transgenic Tobacco

5.1 Introduction 88

5.1.1 Engineered male sterility based on transgenes of non-plant origin 88

5.1.2 Engineered male sterility based on CMS-determining genes 89

5.2 Chapter Aims 91

5.3 Materials and Methods 92

5.3.1 Materials 92

5.3.1.1 Plasmids 92

5.3.1.2 Primers 92

5.3.1.2.1 Primers used in the assembly of the CaMV35S orfC9 construct 92

5.3.1.2.2 Primers used in the assembly of the Ta29 construct 93

5.3.2 Methods 95

5.3.2.1 The preparation of competent *Agrobacterium* cells 95

5.3.2.2 Transformation of *Agrobacterium* by electroporation 95

5.3.2.3 Transformation of *Nicotiana tabacum* 96

5.3.2.4 Extraction of DNA from leaf tissue 96

5.3.2.5 Pollen grain counts 96

5.4 Results 98

5.4.1 Construction of gene fusions between *orfC9* and the mitochondrial targeting sequence of *ß-atp1* under the transcriptional control of the CaMV35S promoter 98

5.4.1.1 Screening of the transgenic plants by PCR 99

5.4.1.2 Northern blot detection of *orfC9* transcripts in To transgenic lines 100

5.4.1.3 RT-PCR detection of *orfC9* transcripts in total RNA extracted from anthers of To transgenic lines 101

5.4.1.4 Northern blot detection of orfC9 expression in total RNA extracted from leaf tissue in the T1 transgenic lines 101
5.4.1.5 Analysis of sterility based on flower morphology, pollen production and seed set 102
5.4.1.6 Analysis of the seedlings phenotype in the different transgenic lines 104
5.4.1.7 Western blot investigation of ORFC9 protein expression in transgenic lines 105
5.4.2 Use of the Ta29 tapetum specific promoter to drive *orfC9* expression 106
5.4.2.1 Screening of the Ta29 transgenic plants by PCR 107
5.4.2.2 Analysis of *orfC9* transcripts in anther tissue of the To transgenic lines by RT-PCR 107
5.4.2.3 Analysis of sterility by flower morphology 108
5.5 Discussion 110

CHAPTER 6

Pollen Development in Fertile and CMS *L. perenne*

6.1 Introduction 114
6.1.1 An overview of the process of pollen development in higher plants 114
6.1.2 The role of the tapetum during pollen development 115
6.1.3 Pollen development and mutations resulting in a male-sterile phenotype 116
6.2 Chapter Aim 118
6.3 Materials and Methods 119
6.3.1 Plant material 119
6.3.2 Isolation of anthers 119
6.3.3 Fixing and embedding the anther tissue in an epoxy araldite resin 119
6.3.4 Light microscopy 120
6.4 Results 121
6.4.1 Pollen development in fertile and CMS *L. perenne* lines 121
6.5 Discussion 124

CHAPTER 7

Sequence analysis of *L. perenne* mitochondrial genes encoding components of the ribosome and the respiratory complexes

7.1 Introduction 127

7.1.1 Molecular phylogeny based on mitochondrial genes in angiosperms 127

7.1.2 An overview of the mitochondrial genome of cereals 129

7.2 Chapter Aims 130

7.3 Materials and Methods 131

7.3.1 Materials 131

7.3.1.1 Plant material 131

7.3.1.2 Primers used in the amplification and sequencing of mitochondrial genes 131

7.3.2 Methods 134

7.3.2.1 Growth of *L. perenne* seedlings 134

7.3.2.2 Preparation of mtDNA from seedling leaf tissue 134

7.3.2.3 Synthesis of cDNA from total RNA 135

7.3.2.4 PCR amplification from mtDNA and cDNA 136

7.3.2.5 Bioinformatic resources utilised 136

7.4 Results and Discussion 137

7.4.1 Nucleotide sequence of mitochondrial genes of *L. perenne* 137

7.4.2 Phylogenetic analysis of mitochondrial gene sequences 138

7.4.3 RNA editing in *L. perenne* mitochondrial genes 142

7.4.3.1 Variation in RNA editing frequency in mt genes of *L. perenne* relative to other monocot species 143

CHAPTER 8

General Discussion and Future Perspectives

8.1 The Dpo- and Rpo-like sequences in *L. perenne* 147

8.2 The association of *orfC9* with CMS in *L. perenne* 148

8.3 Disintegration of the tapetum and pollen abortion in CMS *L. perenne* 149

8.4 *L. perenne* mitochondrial genes as an evolutionary guide 149

References 151

Summary

This study involved molecular investigations of cytoplasmic male sterility (CMS) in perennial ryegrass (*Lolium perenne* L). The CMS line originated as an interspecific hybrid between *L. perenne* (male parent) and *Festuca pratensis* (female parent) that was subsequently backcrossed to *L. perenne* over 9 generations. CMS is a maternally inherited trait in which plants fail to produce functional pollen, often because of rearrangements of the mitochondrial genome that give rise to chimeric genes coding for mitotoxic polypeptides. Chimeric genes have been causally implicated in male sterility in several CMS systems including those of maize, sunflower, radish and petunia. In the case of *L. perenne*, previous work identified a unique 9.6 kb element called CMSi that was present only in the mitochondrial genome of CMS lines. Nucleotide sequence analysis revealed a chimeric gene, *orfC9*, near the leftmost junction between CMSi and mtDNA that was shown to be co-transcribed with the normal upstream *atp9* gene. Thus, in *L. perenne*, *orfC9* represents a potential CMS-determining gene. This hypothesis was investigated by comparing *atp9/orfC9* expression and transcript editing in male-sterile plants and in plants that contained a CMS cytoplasm and a nuclear restorer-of-fertility (*Rf*) gene. However, no differences between CMS and fertility-restored lines were observed. Nevertheless, attempts to transgenically induce a male-sterile phenotype by targeting the ORFC9 protein to mitochondria of *Nicotiana tabacum* resulted in some transgenic lines that exhibited male sterility.

Nucleotide sequence analysis of the 8.2 kb CMSi region located downstream of *orfC9* revealed the presence of three open reading frames coding for a family B-type DNA polymerase (LpDpo) (ORF1), a single subunit RNA polymerase (LpRpo) (ORF2), and a 40.5 kDa predicted protein of unknown function (ORF3). The LpDpo and LpRpo proteins shared significant similarity with DNA and RNA polymerases, respectively, encoded by linear mitochondrial plasmids of plants and fungi and also with integrated linear plasmid-like sequences found in various plant and fungal mitochondrial genomes. The presence of LpDpo, LpRpo and orf3 transcripts was revealed by RT-PCR analysis which further indicated that LpDpo was co-transcribed with orf3.

Further investigation by PCR, and sequence analysis of a series of fertile lines of *L. perenne* (which were, coincidently, infected with a fungal endophyte) detected sequences highly homologous to LpDpo, LpRpo and ORF3. However, Southern blot analysis of fertile lines failed to detect any fragments corresponding to these orfs. This suggests that in fertile lines a CMSi-like element might exist as a low copy number free-living (episomal) plasmid or, alternatively, as an integrant in a low copy number sub-stoichiometric fragment of the *L. perenne* mitochondrial genome. Phylogenetic analysis of the Dpo- and Rpo-like sequences from fertile lines showed they grouped in the clades containing the LpDpo and LpRpo sequences and were more closely related to linear mitochondrial plasmid-like sequences of plants than to those of fungi.

Ten *L. perenne* mitochondrial protein-coding genes and a single rRNA gene were sequenced in order to investigate phylogenetic relationships between *L. perenne* and 7 other species including members of the *Gramineae* whose mitochondrial genomes had been fully sequenced. Unfortunately, because sequences of orthologous genes from genera closely related to *Lolium* (e.g. *Festuca*) were not available, our analyses focused solely on the evolutionary origins of *L. perenne* in relation to distantly-related monocot species. This study revealed a common ancestry between the species *L. perenne* and *T. aestivum* with an estimated divergence time of ca. 34 - 48 Myr ago.

A comparative light microscopic analysis of anthers, at different developmental stages, from fertile and CMS *L. perenne* plants indicated normal pollen development until the post-meiotic tetrad stage. Abnormal vacuolation of the tapetum was observed in CMS anthers following the release of the young microspores from tetrads. This was followed by rapid degeneration of the tapetal cell layer and pollen abortion.

Abbreviations

ABI	Applied Biosystems
Amp	ampicillin
APS	ammonium persulphate
atp	gene encoding an ATPase subunit
BLAST	basic local alignment search tool
bp	base pair
BRU	Bio Resources Unit
BSA	bovine serum albumin
BS	bootstrap
BCL	bootstrap confidence level
°C	degree Celcius
ca.	approximately
$CaCl_2$	calcium chloride
CaMV	cauliflower mosaic virus
Carb	carbenicillin
cDNA	complementary DNA
CMS	cytoplasmic male sterility
cob	gene encoding cytochrome b
cox	gene encoding cytochrome oxidase
cpDNA	chloroplast DNA
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
Dpo	DNA polymerase
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetra acetic acid disodium salt
Exo	3'-5' exonuclease
FIp	fertile *L. perenne* line from Irish pasture
FNz	fertile *L. perenne* line from New Zealand
g	gravitational acceleration/gram
HCl	hydrochloric acid
hr	hour
Ip	insoluble protein
IPTG	isopropyl-ß-D-thiogalactopyranoside
Kan	kanamycin sulphate
Kan R	kanamycin resistant
kb	kilobase
kDa	kilodalton
kV	kilovolt
l	litre
LB	Luria-Bertani
Lp	*Lolium perenne*
µF	microfarad
µg	microgram
µl	microlitre
µM	micromolar
M	molar
mA	milliamp

mCi	milliCurie
MEGA	multiple evolutionary genetic analysis
Met	methionine
mg	milligram
Mg^{2+}	magnesium ion
$MgSO_4$	magnesium sulphate
min	minute
ml	millilitre
mm	millimetre
mM	millimolar
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
mRNA	messenger RNA
mt	mitochondrial
mtDNA	mitochondrial DNA
mw	molecular weight
Myr	Million years
NaCl	Sodium chloride
ng	nanogram
nm	nanometre
nt	nucleotide
OD	optical density
oligo	oligonucleotide
ORF	open reading frame
PBS	phosphate-buffered saline
pga	pollen grains per anther
PCD	programmed cell death
PCR	polymerase chain reaction
Pfu	*Pyrococcus furiosus*
PMC	pollen mother cell
pmol	picomole
PMSF	phenylmethylsulfonyl fluoride
Pol	polymerase
RNA	ribonucleic acid
RNasin	RNAase inhibitor
PPR	pentatricopeptide repeat
rpm	revolutions per minute
rpl	gene encoding a ribosomal protein of the large (50S) subunit
rps	gene encoding a ribosomal protein of the small (30S) subunit
rrn	gene encoding a ribosomal RNA
rRNA	ribosomal RNA
Rpo	RNA polymerase
Rst	restored to fertility
RT	room temperature
RT-PCR	reverse transcription-polymerase chain reaction
SDS-PAGE	sodium dodecyl sulphate-polyacyrlamide gel electrophoresis
ssDNA	single-stranded DNA
Strep	streptomycin sulphate
Ta	tapetum
TP	terminal protein

TB	terrific broth
TBE	Tris-boric acid-EDTA buffer
TBS	Tris-buffered saline
TCD	Trinity College, Dublin
tcDNA	total cellular DNA
TE	Tris-EDTA buffer
trn	gene encoding a tRNA
tRNA	transfer RNA
dH_2O	deionized water

CHAPTER 1

General Introduction

1.1 The Mitochondrial Genome in Plants

Mitochondrial (mt) genomes range in size from the extremely compact 15 - 16 kb found throughout the animal kingdom, the intermediate 20 - 100 kb in protists, fungi and algae, to the extraordinarily diverse 200 - 2,400 kb found in higher plants. The first plant mt genome to be fully sequenced was that of the liverwort *Marchantia polymorpha* (Oda, *et al.*, 1992) with a size of 180 kb. In higher plants, complete mt genome sequences are now available for *Arabidopsis thaliana* (Unseld *et al.*, 1997), *Beta vulgaris* (Kubo *et al.*, 2000), *Brassica napus* (Handa, 2003), *Oryza sativa* (Notsu *et al.*, 2002), *Zea mays* NB (Clifton *et al.*, 2004) and *Triticum aestivum* (Oghihara *et al.*, 2005).

Mitochondrial genome size differences are mainly due to intergenic (i.e. non-coding) sequences and to introns. Important insights into mt genome structure and evolution have recently come from sequence comparisons between *B. napus* and *Arabidopsis* (Handa, 2003). Despite their close phylogenetic relatedness (both are members of the *Brassicaceae*), the *Arabidopsis* mt genome (367 kb) is 65% larger with no increase in coding capacity. Although their encoded proteins show almost 99% sequence similarity, remarkably, 1/3 and 2/3, respectively, of the *B. napus* and *Arabidopsis* mt genomes do not align and show no homology to other plant mt genomes (Handa, 2003). This indicates that after divergence of the two species, the evolution of coding sequences was extremely constrained while the non-coding regions of both mt genomes evolved so rapidly, by a process involving sequence acquisitions and losses, that they became essentially species-specific.

The acquisition of sequences of non-mitochondrial origin can also contribute substantially to mt genome size. These include sequences derived from the chloroplast and nuclear genomes, and sequences derived from linear plasmids.

1.1.1 Gene content of plant mitochondrial genomes – an overview

Mitochondrial genes can be divided into two main classes: those encoding structural RNA components of the translational apparatus (rRNA and tRNA) and those encoding proteins. A comparison of the gene content of animal, fungal and plant mt genomes shows an increase in informational content in plants that includes genes coding for ribosomal proteins, subunits of the respiratory complexes involved in ATP metabolism and *cytochrome c* biogenesis (CCB), and small open reading frames (ORFs) of unknown function (Boore, 1999; Knoop, 2004).

The data presented in Table 1.1 are based on the gene content of the 7 completely sequenced higher plant mt genomes and the liverwort *Marchantia* (Oda *et al.,* 1992). The mt genome of *Marchantia* encodes 94 proteins while the fully sequenced higher plant mt genomes code for 30 – 39, many more than do animals or fungi. With the exception of the genes *atp4*, *atp8* (*orfB*), and *nad9*, *Marchantia* contains the same respiratory genes found in the mitochondrial genomes of the angiosperm species listed in Table 1.1. This includes the six subunits of complex I (*nad1* - *nad6*), Complex III (cytochrome b (*cob*)), three subunits of complex IV (*coxI-coxIII*) and the ATPase subunits 1, 6 and 9 (*atp1*, *atp6* and *atp9*).

Differential loss of several protein coding genes for each plant species is also apparent in Table 1.1. For example, the closely related *Arabidopsis* and *Brassica* differ with respect to other angiosperms and *Marchantia* in that they lack *rps13* and *ccmFN* but contain instead the 2 genes *ccmFN1*, *ccmFN2*, which are absent in *Marchantia* and angiosperms. Table 1.1 also shows the loss of protein coding genes following the divergence of *Marchantia* from a common angiosperm ancestor. Only 3 genes i.e. *rps8, rps11* and *rpl6,* are not found in any of the angiosperms (*rps11* is a pseudogene in rice). The *rpl*6 gene was transferred to the nucleus in *Arabidopsis* (*Arabidopsis* Genome initiative, 2000), while *rps8* was replaced by a duplicate copy of the nuclear gene *rps15A* (Adams *et al.*, 2002).

A recent large-scale survey of 280 flowering plant genera for the presence or absence of 40 genes encoding mt proteins, however, showed extensive variation in mt gene content (Adams *et al.*, 2002). In particular, 16 genes showed patterns of frequent loss, of which, remarkably, 14 encoded ribosomal proteins,

while the remaining 2 code for succinate dehydrogenases. In contrast out of the remaining 24 protein coding genes, only 2 losses were detected.

The number of mt genes encoding tRNAs also differs between plant species, for example, *Marchantia* encodes 29 tRNAs (Oda *et al.*,1992), *Arabidopsis* and maize 22 tRNAs (Unseld *et al.*, 1997) (Table 1.1). Thus, several tRNA genes including *trn*A, *trn*L, *trn*R, *trn*T and *trn*V have been lost following the divergence of the ancestral angiosperms from the liverwort *Marchantia*. Further losses occurred in specific angiosperm lineages. The lineage leading to rice, for example, shows a loss of *trn*G following the divergence of monocot and dicot species (Sugiyama *et al.*, 2005). Similarly, within the dicots, the lineage leading to the common ancestor of *Brassica* and *Arabidopsis* lost the *trn*F gene (Sugiyama *et al.*, 2005). Remarkably, the transfer of several plastid-derived tRNA genes e.g. tRNA His (GUG), Asn (GUU), Pro (UGG), Ser (GGA) and Trp (CCA) to mitochondria was a common event during angiosperm evolution (Sugiyama *et al.*, 2005).

Although most tRNAs are encoded in the mt genome, no higher plant species contains a full complement i.e. sufficient to decode all codons. Hence, 6 – 9 tRNAs, depending on the species, are encoded by nuclear genes and imported from the cytosol (Unseld *et al.*, 1997).

1.1.2 Mitochondrial genes of non-mitochondrial origin

Higher plant mtDNA, including that of *Arabidopsis*, contains so-called promiscuous sequences derived from the plastid or nuclear genome. About 1% of the *Arabidopsis* mt genome comprises sequences of plastid origin which include fragments of photosynthetic genes like *psbD*, *rbcL*, trn*Met* and *ndhB* genes. The fragments vary in size from 30 to 930 nucleotides and are non-functional in the mitochondria but some tRNAs of plastid origin e.g. tRNA *Asp* and tRNA *Ser* are functionally integrated into the mitochondrial translational system (Unseld *et al.*, 1997). In the sugar beet *B. vulgaris,* the plastid-derived sequences are more extensive and comprise 2.1% of the mitochondrial genome. Here, plastid DNA insertions vary in size from 25 bp to 3.4 kb and include a fragment containing the genes *rps7*, *rps12* and tRNAVal-GAC (Kubo *et al.*, 2000). Sequence fragments of nuclear origin are present in mitochondrial DNA of both *Arabidopsis* and *B. vulgaris* which accounts respectively for 4% and 3.3% of the mitochondrial

genome. These fragments are retrotransposon-like and only two of these transposon sequences show similarity between *Arabidopsis* and *B. vulgaris* (Unseld *et al.*, 1997; Kubo *et al.*, 2000).

1.1.3 Replication of plant mitochondrial DNA

Replication of the circular mt genomes of protists, and fungi occurs either by a theta or a sigma (rolling circle) mechanism (Bendich, 1993; Maleszka *et al.*, 1991; Presier *et al.*, 1996), while in organisms such as *Schizosaccharomyces torulopsis* (Sena *et al.*, 1986) and *Plasmodium falciparum* (Preiser *et al.*, 1996) it takes place by a sigma process that involves the formation of circles with linear tails giving rise to a lariat-type structure.

Replication of mtDNA has been closely studied in the dicot plant species *Chenopodium album*. Electron microscopy has revealed the presence of putative intermediates of a rolling circle (RC) type replication giving rise to both branched structures and sigma or lariat-like molecules (Backert *et al.*, 1996a; Backert *et al.*, 1996b; reviewed in Backert *et al.*, 1997b). The replication intermediates revealed 18% of sigma like molecules amongst sub-genomic molecules in cell suspensions. About half of the sigma like molecules had tails that exceeded the contour length of the corresponding putative parental circle and such tails were proposed to represent concatemeric products of rolling circle replication (Backert *et al.*, 1996a). Analysis by 2-dimentional gel electrophoresis lead to the identification of single-stranded DNA circles ranging in sizes from 1.3 kb to 18.7 kb, and the distribution of single- stranded mtDNA in sigma structures indicated a rolling circle-type replication of subgenomic circles (Backert *et al.*, 1997a; Backert *et al.*, 1996b; Backert *et al.*, 1997b). A further study by Backert and Börner (2000) visualized replication and recombination intermediates in mtDNA from *Chenopodium* as branched rosette-like structures. Many of the linear DNA molecules participating in the formation of these branched structures were found to have extended single-stranded ends. Such ends may recombinatively invade homologous regions of double-stranded DNA. Thus, recombination events lead to the formation of complex molecules and initiate replication in a manner similar to that occurring in phage T4. A recombination-dependent mode of replication was therefore proposed to co-exist with a recombination-independent rolling circle mode of replication in the mitochondria of *Chenopodium.*

Remarkably, despite the complete sequencing of the *Arabidopsis*, rice and poplar genomes, knowledge of the mtDNA replication machinery is limited. In fungi and animals, DNA polymerase (Dpo) γ, a member of Dpo family A, has been shown to both repair and replicate mtDNA (Torri and Englund, 1995). However, Dpo activity in nuclear and organelle extracts isolated from tobacco BY-2 cells indicated that plastid and mtDpos (identified as proteins of 116 kDa) were, unlike animal mtDpo γ, relatively insensitive to ddCTP (Sakai *et al.*, 1999). This suggested that Dpo γ might be absent in plants which was subsequently confirmed by genome sequencing. Instead, plants were found to contain homologs of bacterial Dpo I that are predicted to be targeted to organelles. Mori *et al.*, (2005) identified and cloned 2 Arabidopsis Dpo I homologs, *AtPolI-like A* and *AtPolI-like B*, and showed by means of GFP fusions that the former was targeted to plastids and the latter to both plastids and mitochondria. More recently, Ono *et al.* (2007) have shown that the proteins encoded by the tobacco orthologs of the *Arabidopsis* genes are both targeted to mitochondria and plastids and finally demonstrated, via expression in *E. coli,* that the 116 kDa proteins actually possess Dpo I activity.

1.1.4 Recombination and the multipartite structure of plant mitochondrial genomes

Over the past 25 years the scientific literature has extensively documented the recombinogenic nature of the plant mt genome. *Brassica hirta* (208 kb) and *Marchantia* (187 kb) are representative of ancestral forms of the mt genome in that they comprise a single circle (Palmer and Hebron, 1986; Oda *et al.*, 1992). In contrast in most higher plants, the genome exists as subgenomic circles that can recombine to generate a "whole genome" Master Circle (MC). In *B. napus*, the mitochondrial genome exists in the form of three different circular structures: a MC and 2 subgenomic molecules that together comprise the entire mt genome (Palmer and Shields, 1984). In *Arabidopsis*, two pairs of repeated sequences produce a complex population of sub-genomic molecules (Klein *et al.*, 1994) derived from MC. Recent reports have also identified a similar sub-genomic architecture in the monocots *Z. mays* NB (Clifton *et al.*, 2004) and *T. aestivum* (Ogihara *et al.*, 2005). The subgenomic circles are believed to result from recombination events via repeated sequence motifs (usually 1-14 kb in length)

(Fauron *et al.*, 1995) and vary in size and sequence from one plant species to the next, thus emphasising their uniqueness and complexity.

In maize (Fauron *et al.*, 1995; Clifton *et al.*, 2004) and wheat (Ogihara *et al.*, 2005), further complexity is caused by intra-molecular recombination between sub-genomic circles, thus resulting in the generation of further molecular forms and duplication of specific sequences. In general, across the angiosperms recombination via large repeats occurs at high frequency and is reversible, with a state of equilibrium found to exist between the various sub-genomic forms (Palmer and Shields, 1984; Oro *et al.*, 1985). However, dramatic shifts in the stoichiometry of subgenomic molecules have also been documented, the result of a phenomenon termed sub-stoichiometric shifting (SSS) (Small *et al.*, 1989; Abdelnoor *et al.*, 2003). This can dramatically alter the copy number of individual sub-genomic forms, and lead to massive amplification of individual genes or, potentially, of integrated promiscuous DNA sequences. SSS has recently been shown, in *Arabidopsis,* to be under the control of the nuclear gene *Msh1* (Abdelnoor *et al.*, 2003) and homologs of *RecA*. The protein encoded by *Msh1* comprises an N-terminal domain sharing homology with *E. coli MutS* and a C-terminal domain sharing homology with homing endonucleases.

1.1.5 Recombination and mitochondrial mutations

In plants, most mitochondrial mutations have been shown to derive from rearrangement, duplication and deletions of mtDNA loci (reviewed in Newton *et al.*, 2004) due to high frequency reversible recombination events via large repeat sequences. However, they may also be a consequence of (much rarer) irreversible homologous recombination events that occur via short repeats in the genome (typically 6 bp to 100 bp in size). Irreversible recombination events can result in the generation of chimeric orfs or genes. These typically show a mosaic structure with some regions sharing, and other regions lacking sequence homology with the rest of the mt genome. Chimeric genes are not necessarily deleterious to the plant. For example, *Z. mays* (NB) contains 121 ORFs most of which are unique to maize, of which two are found to be truly chimeric: *orf186* contains sequences derived from *coxII,* while *orf127* contains sequences from *rps12* (Clifton *et al.*, 2004). However, in certain instances, particularly among the angiosperms,

chimeric genes can induce a male-sterile phenotype termed cytoplasmic male sterility (CMS) (see below section 1.2).

Several studies have shown that a combination of rare irreversible recombination events between short repeats and frequent recombination events across large repeats can result in deletion of one part and duplication of another part of the mt genome (Small *et al.*, 1989; Fauron *et al.*, 1992). This process has been used to explain the origin of *non-chromosomal stripe* mutants in maize (Newton *et al.*, 1996) and tomato (Bonnema *et al.*, 1995), *maternal distorted leaf* in *Arabidopsis* (Sakamoto *et al.*, 1996) and *mitochondrially-encoded mosaic* in cucumber (Lilly *et al.*, 2001).

1.1.6 Mitochondrial plasmids

The complexity of mtDNA in plants is increased by the presence of plasmids. Two classes of plasmids have been described: circular plasmids which exist exclusively as extrachromosomal replicons in the mitochondrion or linear plasmids that can exist extrachromosomally or can integrate into mtDNA. Small circular plasmid-like DNAs have been detected in the mitochondria of higher plants such as maize (Smith and Pring, 1987), sugar beet (Hansen and Marcker 1984; reviewed in Pring and Lonsdale, 1985), rice (Shikanai and Yamado, 1988), sunflower (de la Canal *et al.*, 1991), broad bean (Wahleithner and Wolstenholm, 1987) and date palm (Benslimane *et al.*, 1994) and the sequence of these plasmids is unrelated to that of the mt genome.

Linear plasmids are defined by terminal inverted repeat (TIR) motifs that share limited sequence homology with the mt genome. Maize plants that carry the CMS-S cytoplasm, for example, contain both extrachromosomal and integrated copies of two linear plasmids, S1 and S2. Integration occurs via homologous recombination between the plasmid TIRs (208 bp) and sequences in the main mt genome (208bp) (Levings and Sederoff, 1983; Paillard *et al.*, 1985). Fragments of linear plasmid-like sequences have been identified in the mt genomes of *B. vulgaris* (Kubo *et al.*, 2000; Satoh *et al.*, 2004), *Daucus carota* (Robison and Wolyn, 2005), *O. sativa* (Notsu *et al.*, 2002) and *Secale cereale* (Dohmen and Tudzynski, 1994). Surprisingly, however, although *B. napus* mitochondria have been shown to contain an extrachromosomal linear plasmid of 11.6 kb, the

completely sequenced mt genome showed no evidence of plasmid integration events (Handa, 2003).

1.1.7 Gene expression in plant mitochondria

Gene expression in plant mitochondria is a complex process at both the transcriptional and post-transcriptional levels. Unlike plastids, mitochondria do not encode a eubacterial multisubunit RNA polymerase (RNAP). Instead, all transcription is mediated by nucleus-encoded, single subunit RNAPs that are phage T7-like (RpoT), and are targeted to the organelle. *Arabidopsis*, and probably all other angiosperms, contains 3 *RpoT* gene classes that code for a mitochondrial RNAP (RpoTm), a plastidial RNAP (RpoTp, also known as NEP) and a polymerase that is targeted to both organelles (RpoTmp) (Hedke *et al.*, 1997, 1999, 2000).

Extensive bioinformatics-based sequence analyses have been used to identify consensus promoter sequence elements in plant mt genomes. Many such promoters contain a conserved ATTA, RRTA or YRTA motif (where Y = pyrimidine and R = purine) located immediately upstream of the transcription start site (summarized in Fey and Marechal-Drouard, 1999) and so resemble plastid promoters recognized by RpoTp. Recently, using recombinant RpoTm, RpoTp and RpoTmp proteins produced in *E. coli*, Kuhn *et al.*, (2007) have shown that RpoTm and RpoTp accurately initiate transcription *in vitro* from mitochondrial and plastid promoters, respectively, without the aid of protein co-factors. However, although RpoTmp was transcriptionally active, it did not show promoter specificity. Thus, RpoTm is likely to be the major RNAP responsible for mt gene transcription in *Arabidopsis*.

Gene expression in mitochondria is also profoundly affected by post-transcriptional processes that include RNA editing, cis- and trans- splicing, RNA trimming, polyadenylation and the regulation of transcript stability. RNA editing, in particular, plays a key role in post-transcriptional regulation of mt gene expression, and over the last decade has attracted considerable attention (reviewed most recently in Shikanai, 2006).

1.1.7.1 RNA editing in plant mitochondria

The term "editing" was originally applied to RNA transcripts to explain the addition and deletion of uridine (U) residues that affect mRNA transcripts in trypanosomes (Benne *et al.*, 1986). It was later used to explain the post-transcriptional conversion of C to U residues (or the rarer conversion of U to C), that results in an RNA with a base sequence that differs from the DNA molecule encoding it. In practice, therefore, RNA editing sites can be identified by comparing the sequence of a cDNA with its corresponding genomic sequence.

RNA editing is more common in plant mitochondria than in chloroplasts. The mitochondrial mRNAs of *Arabidopsis* and rice are edited at 441 and 491 sites (C to U) respectively, compared with 19 and 21 editing sites respectively, in chloroplast mRNAs. None of the rarer U to C transitions have been detected in either chloroplasts and mitochondria of *Arabidopsis*, or in mitochondria of rice, but such changes are known to be quite common in mosses and ferns (Malek *et al.*, 1996). In some species such as the hornworts, more than one third of the editing sites are U to C transitions which function to eliminate genomic in-frame stop codons (TGA or TAA). In flowering plants there have only been a few reports of U to C editing: *coxIII* in wheat (Gualberto *et al.*, 1990), *cob* in *Oenothera* (Schuster *et al.*, 1990).

To date RNA editing has been assessed for most of the known protein genes in the mt genomes of the angiosperms *Arabidopsis* (Geige and Brennicke, 1999), *B. vulgaris* (Kubo *et al.*, 2000), *B. napus* (Handa, 2003) and *O. sativa* (Notsu *et al.*, 2002). In *Arabidopsis* mitochondria 8.0% of all C-containing codons are edited but the editing frequencies are distributed unevenly between different genes (Giege and Brennicke, 1999). For example the *coxI* transcript is not edited at all but the *ccb2* mRNA is altered at 39 sites. Furthermore, a comparison between gene groups showed that mRNAs encoding CCB and complex I proteins are edited at higher frequencies than other gene groups. In a further analysis of RNA editing in *Arabidopsis* 35% of the modifications were shown to result in hydrophilic to hydrophobic amino acid substitutions (Giege and Brennicke, 1999). Confirmation of a key role played by the editing process in protein functionality has been shown by targeting a variant form of the ATP9 protein, encoded by an unedited transcript, into the mitochondria of tobacco. Transgenic plants were male sterile, indicating that assembly of unedited ATP9

protein in the mitochondria likely affected the production of energy required during a critical stage of pollen development (Hernould *et al.*, 1993).

Editing has also been identified in one class of structural RNA i.e. tRNA (Fey *et al.*, 2002), with no reports of editing in the three rRNAs (5S, 18S and 26S (Giege and Brennicke, 1999; Gagliardi and Gualberto, 2004). Functionally important RNA editing has been found in tRNAPhe and tRNACys from *Oenothera* (Binder *et al.*, 1994), whereas no editing has been identified in any of the cDNA clones from the tRNAs in *Arabidopsis* (Giege and Brennicke, 1999). *In vitro* studies of edited and non-edited tRNAPhe precursor molecules showed that editing is a prerequisite for 5' and 3' processing and as a result editing can have an effect on the amount of mature tRNAPhe present in the total tRNA population (Marechal- Drouard *et al.*, 1996).

1.1.7.2 The mechanism of RNA editing site selection

The question as to how a given C in an RNA molecule in plant mitochondria is identified for conversion to a U remains unanswered. Both *in vivo* and *in vitro* methods have been used to identify *cis*-specificity elements in plastid and mitochondrial mRNAs. These investigations have shown that sequences 5' of the edited nucleotide are the main determinants in site recognition. (Chaudhuri *et al.*, 1997; Farré *et al.*, 2001; Miyamoto *et al.*, 2004). In pea, *in vitro* RNA editing of *atp9* identified a critical *cis*-element region extending 20-30 bases upstream of a C edited site (Takenaka *et al.*, 2004). Most recently, an extended upstream *cis*-element of 30/40 bases was shown to be required for *atp9* editing in cauliflower (Neuwirt *et al.*, 2005), suggesting the evolution of species-specific editing site *cis*-elements, at least in the case of *atp9* editing.

Editing of organelle transcripts is also dependent on proteins encoded by nuclear genes. Pentatricopeptide repeat (PPR) proteins, which have been shown to bind to specific mRNAs, and to be involved in RNA maturation processes such as RNA stabilization, RNA cleavage and splicing are likely candidates (Lurin *et al.*, 2004). A PPR protein has recently been identified as a likely *trans*-acting factor involved in RNA editing in chloroplasts (Kotera *et al.*, 2005), and several nuclear genes that act to restore fertility in plants showing cytoplasmic male sterility have recently been shown to encode a mitochondrion-targeted PPR protein (Bentolila *et al.*, 2002; Brown *et al.*, 2003; Klein *et al.*, 2005; Wang *et al.*, 2006). Their

involvement in organelle-related processes that include RNA editing may account for the multiplicity of PPR proteins in plants. The genome of *Arabidopsis*, for example, encodes some 450 PPR proteins.

1.2 Cytoplasmic Male Sterility in Flowering Plants

Mutations that interfere with pollen production cause male sterility in plants. The sterility may be due to mutations in nuclear genes in which case it shows Mendelian inheritance and is referred to as Nuclear Male Sterility, or alternatively, to mutations in mitochondrial genes in which case it shows non-Mendelian (usually maternal) inheritance and is referred to as Cytoplasmic Male Sterility (CMS).

CMS is the most intensively investigated (and used) male-sterility system in plants. It is a two-component system comprising a maternally-inherited mutant mitochondrial gene that disrupts pollen development and a nuclear gene (*Rf*, for restorer of fertility) that complements (or suppresses) the mitochondrial defect, thus restoring male fertility. A remarkable feature of most CMS systems is the developmental- and tissue-specificity of mutant phenotype expression: while CMS and wild-type fertile plants show similar vegetative growth and development rates, the former do not produce viable pollen. To date CMS has been identified in well over 150 plant species (Schnable and Wise, 1998).

CMS has proved to be especially useful in the production of F1 hybrid seed, which as a result of heterozygote superiority (or heterosis), gives rise to plants showing increased vigour, disease resistance and improved yield. For relatively small-scale horticultural crops such as tomato and cucumber, hybrid seed production is generally accomplished by hand emasculation of the female parent (in order to prevent self-pollination), followed by cross-fertilization with pollen of the male parent. However, for large-scale commercial production of hybrid seed, for seed crops such as maize or forage crops such as *Lolium perenne* the perennial ryegrass, a biological system i.e. CMS, that renders the female parent male-sterile, but facilitates restoration of fertility in plants grown from hybrid seed, is highly desirable.

A dramatic expansion in maize hybrid seed production was facilitated in the 1950s by the introduction of maize lines carrying the Texas male-sterile cytoplasm (CMS-T). This eliminated the need for the labour-intensive hand-emasculation of female parental lines necessary to prevent self-fertilization. The female lines carried the CMS-T cytoplasm but not the nuclear restorer of fertility (*Rf*) genes, which were present in the male parental lines. Cross-pollination results in F1 hybrid seeds carrying the CMS cytoplasm but yielding fertile plants due to the activity of the *Rf* genes. Such was the success of this strategy that by the 1970s 85% of US maize carried the CMS-T cytoplasm. However, unexpectedly CMS-T maize proved to be especially susceptible to Southern corn leaf blight caused by race T of the fungal pathogen *Bipolaris maydis.* The wet spring of 1970 provided optimal conditions for fungal growth resulting in the destruction of most of the maize crop in the Southern states. The devastation caused by the CMS-T and its specific interaction with *B. maydis* race T, ultimately led to the abandonment of CMS strategies for hybrid seed production in maize (Levings, 1993).

CMS systems have also been identified and investigated in sugar beet, sunflower, petunia, *B. napus*, rice, sorghum and wheat. CMS can arise naturally as a result of rare spontaneous mutations or as part of a plant breeding programme. The best example of the latter is the *PET1* CMS-inducing cytoplasm in sunflower (*Helianthus* sp.), which derived from an interspecific cross between *Helianthus petiolaris* and *H. annuus* (Horn and Friedt, 1999; Balk and Leaver, 2001).

1.2.1 Mitochondrial genes associated with cytoplasmic male sterility

At the molecular level, many investigations have demonstrated an association between novel or chimeric mitochondrial genes and CMS. Different methods have been applied to assist in the detection of CMS-determining genes (CMS genes), such as comparative restriction fragment length polymorphism (RFLP) analyses of mtDNA in fertile and male-sterile lines. Most CMS genes comprise chimeric open reading frames (*orfs*) that are often the product of a series of complex mitochondrial DNA rearrangements and consist of fragments of known mitochondrial genes fused with sequences of unknown origin. One method to establish a link between the CMS trait and a given mitochondrial locus

is to compare the expression of these orfs in male-sterile and restored fertile lines, as shown for *B. vulgaris* and *B. oleracea* (Satoh *et al.*, 2004; Shinada *et al.*, 2006). Unfortunately, it is not yet possible to stably transform plant mitochondria, and transformation of the nuclear genome with a transgene construct that targets candidate CMS-determining proteins (CMS proteins) into mitochondria has only occasionally proven successful in producing transgenic plants with a male-sterile phenotype (e.g. Wang *et al.*, 2006).

In flowering plants to date ca. 13 mtDNA regions have been associated with CMS. Remarkably, the chimeric orfs they contain have in many cases been shown to contain coding sequences derived from genes encoding subunits of the ATP synthase complex or Complex IV (cytochrome oxidase). In cases where the chimeric orf does not contain *atp* gene sequences, the orf is often located near an *atp* gene.

1.2.2 Chimeric genes and CMS

The majority of plants that carry CMS orfs contain a full complement of wild-type mt respiratory genes. Hence, CMS orfs represent dominant 'gain-of-function' mutations. The orfs are usually generated as a result of recombination events that favour the placement of the orf downstream from sequences supporting their expression. A common finding is that CMS orfs are fused directly to plant mitochondrial promoter sequences or are co-transcribed with upstream mitochondrial respiratory complex genes (Hanson and Bentolila, 2004).

The majority of CMS orfs are unique both in terms of their nucleotide and amino acid sequence, but there are two exceptions to these findings. In rice, the C-terminus of the Boro II CMS-associated protein ORF79 shows 80% sequence identity with the predicted protein of a CMS candidate locus *orf107* in *Sorghum* A3 (Tang *et al.*, 1996), thus illustrating an unusual situation of sequence similarity between CMS-determining loci in the mitochondria of two different plant species.

This suggested that the C-terminal region of these polypeptides might be important in the detrimental effect their expression has on pollen development. This was supported by the observation that *orf79* encodes a transmembrane protein that behaves as a cytotoxic polypeptide in *E. coli*, causing a decrease in cell density and lethality which depends on the presence of a 5 amino acid

segment of the C-terminal region (Wang *et al.*, 2006). Transformation of normal fertile rice with a nuclear *orf79* transgene was found to result in 50% male sterility, thus confirming that *orf79* encodes a protein that is a causative agent in CMS in rice. Immunoblot analysis confirmed that the ORF79 protein accumulated specifically in microspores of the CMS line but was absent in vegetative tissue and in microspores of lines restored to fertility. This supports the idea that post-transcriptional or post-translational regulation may suppress the accumulation of the ORF79 protein in vegetative tissue and restored anthers (Wang *et al.*, 2006).

Similar observations of cytotoxicity have been made for the maize CMS-T gene T-*urf13* (Dewey *et al.*, 1987) and the CMS genes *orf552* in *Helianthus* and *orf138* in *Brassica* each of which proved to be toxic when expressed in *E.coli* (Nakai *et al.*, 1995; Duroc *et al.*, 2005). T-URF13 has been shown to exist as an oligomer that spans the inner mitochondrial membrane where it functions as a ligand-gated pore in response to T-toxin (produced by *B. maydis* race T) or the insecticide methomyl that facilitates ion leakage and the uncoupling of oxidative phosphorylation (Levings, 1990).

Taken together, the various findings discussed above suggest that CMS-determining genes may disrupt pollen development because they encode proteins that act primarily to interfere with mitochondrial function.

1.2.2.1 How does CMS affect male fertility but, in general, not vegetative growth?

The exact mechanism by which CMS proteins lead to male sterility has not been fully determined for any plant species. A major technical drawback in comparing reproductive tissues from CMS and fertile lines is obtaining sufficient quantities of reproductive tissue at the relevant developmental stage for both molecular studies and physiological and biochemical analyses. For example, in maize, mitochondrial disruption leading to CMS occurs early in pollen development (Warmke and Lee, 1978) when the anthers are small, thus making it difficult to obtain sufficient quantities of tissue for biochemical analyses (Hanson and Bentolila, 2004). For this reason, and the fact that CMS-determining proteins are expressed in all tissue, most of the investigations performed to date have tended to focus on vegetative tissues.

Perhaps the most intriguing question in any CMS system concerns the issue of specificity: given its constitutive expression, how does the CMS protein act to specifically disrupt pollen development without, apparently, affecting other aspects of plant growth and development? In the common bean *Phaseolus vulgaris* the reason for this appears to be straightforward; the CMS protein ORF239 is associated with the cell wall of developing microspores and is degraded by proteases in vegetative tissue (Abad *et al.*, 1995; Sarria *et al.*, 1998). Transgenic plants expressing a nuclear transgene encoding ORF239 were found to be male-sterile (He *et al.*, 1996), thus showing that the ORF239 protein in this instance is toxic to developing pollen grains whether it be encoded in the mitochondrion or in the nucleus.

The so-called 'energy hypothesis' was advanced to explain the specific disruption of pollen development in maize CMS lines (Warmke and Lee, 1978; Hanson and Bentolia, 2004; Teixeira *et al.*, 2005b). This is based on the observation that the normal development of tapetal cells and pollen is accompanied by a 20 - 40 fold increase in the number of mitochondria per cell and assumes that the increase reflects an increased energy demand. The hypothesis proposes that CMS proteins interfere with mitochondrial energy production which, in consequence, leads to an energy deficit and ultimately, cell death. This seems plausible since some CMS proteins contain domains derived from ATP synthase subunits and the synthesis of several CMS proteins is linked to the synthesis of ATP synthases. Thus, the CMS protein may act to poison the multi-subunit mitochondrial ATP synthase complex and dramatically reduce ATP biosynthetic capacity. Evidence supporting the hypothesis has come from studies of sunflower seedlings which showed that F_0-F_1 ATP synthase activity is reduced in CMS lines compared to fertile restored line (Sabar *et al.*, 2003) and from alloplasmic CMS *Nicotiana* lines, in which the ATP:ADP ratio in floral buds was shown to be lower than in normal fertile tobacco lines (Bergman *et al.*, 2000). These observations suggest that impaired ATP synthase activity is a likely 'casual factor' in the disruption of pollen development in some CMS lines.

An alternative hypothesis advanced by Flavell (1974) proposes that CMS acts specifically to interfere with pollen development due to the presence of an anther specificity factor that is absent in other tissues. In the case of CMS-T, such a factor might for example be a chemical analog of T-toxin which fortuitously

acts as a ligand that opens the T-URF13 pore thereby interfering with mitochondrial energy production. However, anther-specific factors implicated in CMS have not, so far, been identified.

1.2.3 CMS and programmed cell death

A clear association between mitochondrial function and programmed cell death (PCD) has been extensively documented in animal cells (Bras *et al.*, 2005), whereas the role of mitochondria in PCD in plants has not been so well defined. The hallmarks of animal PCD are an early release of *cytochrome c* from mitochondria to the cytosol, followed by a characteristic condensation and fragmentation of the nucleus detected as oligonucleosomal cleavage of DNA. In plants, PCD of anther tissue is ultimately required for the release of mature pollen grains (Papini *et al.*, 1999), suggesting that male sterility may simply reflect an inappropriately early PCD, induced by CMS proteins. This hypothesis was investigated by Balk and Leaver (2001) using sunflower *PET1-CMS* lines. A comparative cytological investigation of anther development in fertile and CMS plants indicated that microspore death in this system was not due to necrosis but was associated with oligonucleosomal cleavage of DNA, a characteristic feature of PCD that had been previously observed in normal barley (*Hordeum vulgare* L.) anthers at pollen maturation (Wang *et al.*, 1999). Balk and Leaver (2001) observed *cytochrome c* release during the development of both fertile and CMS anthers although the quantities released differed at different stages of development. In sterile anthers it occurred during early meiosis when the first signs of abnormal development can be seen, whereas in fertile anthers it occurred later, at the tetrad stage. However, *cytochrome c* release was not observed in all tissues showing features of PCD. Thus the significance of *cytchrome c* release and PCD in CMS remains unresolved.

However there is some evidence that the occurrence of premature PCD might be a key mechanism responsible for CMS (Pelletier and Budar, 2007). This observation states that although the male organ itself is well-preserved, the tapetal layer surrounding the pollen cells is prematurely destroyed in male-sterile plants (Balk and Leaver, 2001). Under normal circumstances in male-fertile organs the tapetal cells undergo a precise pathway of development that apparently involves

appropriately timed PCD (Rogers, 2006). Premature triggering of this pathway may lead to male sterility as was shown in rice (Ku *et al.*, 2003).

1.3 Perennial Ryegrass

The cool season perennial rye grass species *Lolium perenne* L. is best adapted to temperate climates with mild winters and cool moist summers. It was originally recognised as a pasture grass in Europe, and later on was cultivated in other parts of the world. However because of its poor heat and cold tolerance it performed poorly in North America, showing susceptibility to diseases such as 'gray leaf spot' and 'stem rust' in warm and humid regions (reviewed in Bonos *et al.*, 2006). In European temperate climates it is primarily used as a forage and turf grass crop, providing high yields and high nutritive value with good dry matter content and digestibility for grazing animals.

L. perenne is a diploid species (2n=14) and in nature there are no naturally occurring polyploids. Tetraploids (2n=28), which were first developed in the 1960s, differ from diploids in terms of tiller density, water soluble carbohydrate content, digestibility, and greater preference by grazing animals (Connolly *et al.*, 1977; Castle and Watson, 1971). In Ireland the usage of tetraploids is the highest in Europe accounting for 35-40% of seed sold, and as part of a breeding programme the selection of diploids and tetraploids is done in alternate years. The main objective at both the diploid and tetraploid level is to increase annual yield, seasonal yield in spring and autumn, persistency, and resistance to the fungi causing mildew, drechsclera leaf spot and leaf rust.

The development of improved *L. perenne* cultivars has increased steadily worldwide (Huff *et al.*, 1997) due to major breeding programmes in Europe (e.g. Teagasc, Carlow (Ireland)), in the US (Lexington, Kentuckey), and in New Zealand (Ministry of Agriculture, Lincoln). Cultivar improvement has, however, been based on a limited germplasm, with the result that cultivars are somewhat genotypically and phenotypically similar (Funk *et al.*, 1993). Also, *L. perenne* is an outbreeding perennial species with significant self-incompatability potential, thus making each cultivar line a heterogenous mix of genotypes. Despite the challenges this poses for cultivar evaluation, cultivars showing significant

improvements in growth, yield and persistency have been identified and brought to market (Camlin, 1997).

1.3.1 The Ryegrass-Fescue *(Lolium-Festuca)* complex

Lolium and *Festuca* are closely related genera within the family *Poaceae* which contains ca. 10,000 species. Phylogenetic analyses based on RFLPs of non-coding chloroplast DNA sequences and the internal transcribed spacer of the nuclear rRNA genes (Charmet *et al.*, 1997) indicated divergence of the sub-family *Festuceae* from *Poa trivialis* 13 Myr ago. The *Festuceae* then split into broad-leaved (which includes *F. pratensis)* and fine-leaved fescues about 9 Myr ago, and within the broad-leaved group, the genus *Lolium* emerged ca. 2 Myr ago. The emergence of the different *Lolium* species is a comparatively recent event occurring ca. 1 Myr ago.

Species in the genus *Festuca* exhibit several traits that are superior to those found in *Lolium* e.g. greater persistence, salt tolerance, winter hardiness and drought resistance (Thomas and Humphreys, 1991; Thomas *et al.*, 2003). Fortunately, because of the phylogenetic closeness of *Lolium* and *Festuca*, and despite differences in nuclear DNA content, plant breeders have been able to produce fertile *Lolium/Festuca* hybrids with relative ease. The *Lolium* and *Festuca* chromosomes show frequent pairing and recombination in the hybrids, with little or no selection against chromosomes of one species over the other (Zwierzykowski *et al.*, 1998). Also, as expected, the hybrids exhibit large variation in characteristics such as cold tolerance, yield during the summer growth period, and qualities appealing to turf grass developers such as increased vigor and disease resistance, a deeper green colour and increased leaf width and length (Wit, 1974; Kopecky *et al.*, 2005).

1.3.2 CMS in perennial ryegrass

Male-sterility was identified by Connolly and Wright-Turner (1984) in the progeny of a cross between *F. pratensis* (as the female parent) and *L. perenne* (as the male parent). From the original cross, four F1 hybrids seedlings were selected for backcrossing to *L. perenne*. In the early backcross generations the number of sterile genotypes decreased from the second backcross onwards, possibly due to the introduction of fertility restoring genes by some of the pollinators. To resolve

this problem, backcross generations were selected and crossed with a stable maintainer, A19, as the pollen parent. This indicated that the male sterile phenotype was being transmitted through the cytoplasm. An earlier study by Wit (1974) reported the occurrence of CMS in *L. perenne*, but the lack of stable isogenic maintainer lines prevented further analysis and its development for hybrid seed production. Connolly and Wright-Turner (1984) produced inbred maintainers from self-fertilized A19 plants and simultaneously maintained the CMS lines using these inbred lines. In this way the CMS lines from backcross generations 6/7 onwards were genetically identical or isogenic with the appropriate maintainer line. Fertility restoring lines were also identified and examination of the segregation pattern indicated that at least two loci were involved in restoration, and that a single dominant gene at either locus would restore fertility. Moreover, the cytoplasmic nature of male sterility in *L. perenne* was confirmed by reciprocal crosses, a standard test for maternally inherited characters (Kiang *et al.*, 1993). The stability of three CMS lines was tested along with maintainer lines by propagating over 3 generations.

Restriction digestion profiling of mtDNA using the enzyme *Hind*III identified a single band difference between fertile and CMS lines i.e. a 5.6 kb fragment that was present in the former and absent in the latter. A Southern blot analysis using this fragment as a probe led to the identification of a unique 4.5 kb *Bam*HI/*Hind*III fragment in CMS lines which was cloned to give pCMS45. The pCMS45 probe proved to be diagnostic for lines carrying the CMS cytoplasm, thus providing a genetic identifier for the sterile cytoplasm (Kiang *et al.*, 1993). The organelle inheritance pattern was examined by Kiang *et al.* (1994) in the CMS line CMS9B290, a ninth generation backcross line (Connolly and Wright-Turner, 1984). Surprisingly, the restriction enzyme profile of mtDNA from CMS9B290 was more similar to that of the paternal parent (*L. perenne*) than to the maternal parent, *F. pratensis.* However, the paternally inherited mtDNA was not identical to that found in fertile *L. perenne* because of the rearrangement associated with the male sterility phenotype (Kiang *et al.*, 1994; Kiang and Kavanagh, 1996). This was initially investigated by comparative restriction mapping of the CMS-associated mtDNA region from which pCMS45 was derived and the corresponding region in the isogenic fertile maintainer line LPS21. These investigations revealed a rearranged region in CMS mtDNA extending 9.6 kb

relative to fertile mtDNA and the presence of a chimeric gene, *orfC9*, the first 6 codons of which were identical to that of *atp9*. The *orfC9* gene was also shown to be unique to the CMS line and to be co-transcribed with *atp9*. However, its involvement in the CMS phenotype and the origin of the rearranged region was not investigated.

1.4 Thesis Aims

The overall aim of this study was to investigate the origin and role of the LpCMSi region in *L. perenne*, with an initial aim of investigating expression of the CMS-associated *orfC9* in sterile lines in order to establish if it played any role in male sterility (Chapter 3). A further aim was to sequence the entire LpCMSi region with a view to identifying genes that might reveal its evolutionary origins (Chapter 4). The potential use of *orfC9* to induce a male-sterile phenotype in transgenic tobacco plants was also investigated using a nuclear transgene coding for a mitochondrially-targeted ORFC9 protein (Chapter 5). In addition, in order to establish the timing of pollen abortion in *L. perenne* a light microscopic analysis of anther/pollen development in fertile and CMS lines was carried out (Chapter 6). Finally, ten *L. perenne* mitochondrial genes were sequenced and used to investigate phylogenetic relationships with other cereal monocot species (Chapter 7).

Table 1.1 Gene content in several fully sequenced plant mitochondrial genomes

MaizeNB-*Zea mays* (*Zm*), Rice-*Oryza sativa* (*Os*), Wheat-*Triticum aestivum* (*Ta*), Sugar beet-*Beta vulgaris* (*Bv*), *Arabidopsis thaliana* (*At*), Rapeseed-*Brassica napus* (*Bn*), Tobacco-*Nicotiana tabacum* (*Nt*), Liverwort-*Marchantia polymorpha* (*Mp*).

Gene	*Zm*	*Os*	*Ta*	*Nt*	*Bv*	*At*	*Bn*	*Mp*
ComplexI								
nad1	+	+	+	+	+	+	+	+
nad2	+	+	+	+	+	+	+	+
nad3	+	+	+	+	+	+	+	+
nad4	+	+	+	+	+	+	+	+
nad5	+	+	+	+	+	+	+	+
nad6	+	+	+	+	+	+	+	+
nad7	+	+	+	+	+	+	+	Ψ
nad9	+	+	+	+	+	+	+	-
ComplexII								
sdh2	-	-	-	-	-	-	-	+
sdh3	-	-	-	+	-	-	-	+
sdh4	-	-	-	+	-	Ψ	Ψ	+
ComplexIII								
Cob	+	+	+	+	+	+	+	+
ComplexIV								
CoxI	+	+	+	+	+	+	+	+
CoxII	+	+	+	+	+	+	+(2)	+
CoxIII	+	+	+	+	+	+	+	+
Complex V								
atp1	+(2)	+	+	+	+	+	+	+
atp4 (orf25)	+	+	+	+	+	+	+	-
atp6	+	+	+(2)	+	+	+(2)	+	+
atp8 (orfB)	+	+	+(2)	+	+	+	+	-
atp9	+	+	+	+	+	+	+	+
Cytochrome C								
ccmB(206)	+	+	+	+	+	+	+	+
ccmC(250)	+	+	+	+	Ψ	+	+	+
ccmFN(574)	+	+	+	+	+	-	-	+
ccm-FN1(382)	-	-	-	-	-	+	+	-
ccm-FN2(203)	-	-	-	-	-	+	+	-
ccmFC(438)	+	+	+	+	+	+	+	-
Ribosomal Proteins								
rps1	+	+	+	+	-	-	-	+
rps2A	+	+	+	-	-	-	-	+
rps2B	+	-	-	-	-	-	-	-
rps3	+	+	+	+	+	+	+	+
rps4	+	+	+	+	+	+	+	+
rps7	+	+	+	-	+	+	+	+
rps8	-	-	-	-	-	-	-	+
rps10	-	-	-	+	-	-	-	+
rps11	-	Ψ	-	-	-	-	-	+
rps12	+	+	+	+	+	+	+	+
rps13	+	+	+	+	+	-	-	+
rps14	-	Ψ	-	Ψ	-	Ψ	+	+
rps19	-	+	+	+	-	Ψ	-	+
rpl2	-	+	Ψ	+	-	+	+	+
rpl5	-	+	+	+	+	+	+	+
rpl6	-	-	-	-	-	-	-	+
rpl16	+	+	+	+	-	+	+	+

Gene	*Zm*	*Os*	*Ta*	*Nt*	*Bv*	*At*	*Bn*	*Mp*
Other Proteins								
mat-r	+	+	+	+	+	+	+	-
mttB(orfX)	+	+	+	+	+	+	+	-
Ribosomal RNAs								
*rrn*5	+	+	+(3)	+	+	+	+	+
*rrn*18	+	+	+(3)	+	+	+	+	+
*rrn*26	+	+	+(2)	+	+(3)	+	+	+
Transfer RNAs								
*trn*A	Ψ	-	+	-	-	-	-	+
*trn*R	Ψ	Ψ	-	-	-	-	-	+(3)
*trn*N	+(2)	+	+	+	+	+	+	+
*trn*D	+(2)	+	+(2)	+	+	+	+	+
*trn*C	+	+	+	+	+(2)	+	+	+
*trn*E	+(2)	+(2)	+	+(2)	+	+	+	+
*trn*Q	+	+(2)	+(3)	+	+	+	+	+
*trn*G	-	-	-	+	+	+	+	+(2)
*trn*H	+	+(2)	-	+	+	+	+	+
*trn*I	+(3)	+	+	+(2)	+	+	+	+
*trn*K	+	+	+(3)	+	+	+(2)	+	+
*trn*L	Ψ	-	-	-	-	-	-	+(3)
*trn*Me	+	+(3)	+	+	+	+(2)	+	+
*trnf*M	+	+	+(3)	+	+(4)	+	+	+(2)
*trn*F	+	+(2)	+	+	+	-	-	+
*trn*P	+(2)	+(2)	+(2)	+(2)	+	+	+	+
*trn*S	+(2)	+(3)	+(3)	+(3)	+(3)	+(5)	+(2)	+(2)
*trn*T	-	-	-	-	-	-	-	+
*trn*W	+	+(2)	+	+	+	+	+	+
*trn*Y	+	+(2)	+	+	+	+(2)	+	+(2)
*trn*V	Ψ	Ψ	-	-	Ψ	-	-	+

+ gene present; - gene absent; Ψ pseudogene; the numbers in parentheses represent the gene copy number.

CHAPTER 2

General Materials and Methods

2.1 Materials

2.1.1 General chemicals

BDH Ltd, UK:	Chloroform, Ethanol, Formaldehyde solution (37%), Ferric (II) Sulphate, Glycerol, Hydrochloric acid, Iso-propanol, Manganese (II) chloride, Magnesium chloride, Phenol, Sodium bicarbonate, Sodium chloride, Sodium hydrogen phosphate, Sodium di-hydrogen phosphate, Sodium hydroxide, Urea
Invitrogen, UK:	Agarose electrophoresis grade
EC National Diagnostics:	Acrylamide-Urea Concentrate, Diluent, 10X Buffer
Oxoid, UK:	Tryptone, Yeast Extract
Beckton Dickenson, UK:	BBL Grade A agar
Sigma-Aldrich, UK:	Ammonium persulphate (APS), Ampicillin, Bovine serum albumin V, Bromophenol blue, Dextran blue, Diethylpyrocarbonate (DEPC), Ethidium bromide, Ethylenediaminetetraacetic acid disodium salt (EDTA), Formamide, Sodium deocyl sulphate (SDS), Trizma base, Tri-ethylmethylenediamine (TEMED), Triton X-100.

2.1.2 Plant cell culture reagents

Sigma-Aldrich, UK:	Murashige and Skoog Basal Salts, Pyridoxine, Naphthalene acetic acid, Inositol, Thiamine,

	Kanamycin sulphate, Carbenicillin, Streptomycin sulphate.
Irish Merchants, ROI:	Tissue culture pots
Sterilin, UK:	Petri dishes (90mm), 30ml plastic universals, 5 ml Bijoux vials.

2.1.3 Molecular biology

2.1.3.1 Cloning reagents

NEB, UK:	*BamH*I (10 U/µl), *EcoR*I (10 U/µl), *Hind*III (10 U/µl), *Kpn*I (10 U/µl), *Nde*I (10 U/µl), *Sma*I (10 U/µl), *Sal*I (10 U/µl), T4 DNA Ligase (1U/µl)
Stratagene, UK:	pBluescript ® II(+) DNA vector, pUC19 DNA vector, *Pfu* DNA Polymerase 10U/µl
Invitrogen, UK	Isopropyl-ß-D-thiogalactoside (IPTG), 5-bromo-4-chloro-3-indoyl-ß-D-galactoside (X-gal). *Escherichia coli* DH5α Competent cells (Genotype: F^{φ} 80*lacZ*ΔM15Δ (lacZYA-argF)U169*rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) phoA sup E44 thi-1*gyr*A96*rel*A1*ton*A). TOP0-TA® Cloning Kit, TOPO-10 ® Competent cells [Genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ^-] Jetspin PCR kit (Genomed), Jetspin DNA purification kit (Genomed), Jetspin plasmid mini-prep. (Genomed)

2.1.3.2 Radiolabelling

Amersham-GE Health, UK:	[α-^{32}P] dCTP (6000 Ci/mmol), Cycle Sequencing kit RPN2436, Hybond N membrane RPN213, Rediprime labelling kit.

Fuji, UK	X-ray film RX-100

2.1.3.3 PCR amplification (Taq)

Invitrogen, UK:	PCR Buffer (10X), Platinum *Taq* DNA polymerase (5 U/µl), 50 mM $MgCl_2$
Sigma-Aldrich, UK:	dNTPs (100 mM)

2.1.3.4 Reverse transcription (RT)-PCR

Invitrogen, UK:	Superscript ^{TM} II Moloney Murine Leukaemia Virus Reverse Transcriptase. First strand buffer (5X), DTT (0.1M), Random hexamers (50 ng/µl). DNase I enzyme (1 U/µl)
Sigma-Aldrich, UK:	Deoxynucleotidetriphosphates (dNTPs) 100 mM
Promega, UK:	RNasin (40 U/µl)

2.1.3.5 Protein expression analysis

Sample preparation and concentration determination

Sigma-Aldrich, UK:	ß-mercaptoethanol, Glycerol, Sodium dodecyl sulphate (SDS), Tris-Hydrochloride, Bromophenol blue, Phenylmethylsulfonyl fluoride (PMSF), Lysozyme
BioRad, USA:	Bradford reagent

Polyacrylamide Gels and Western blotting

Sigma-Aldrich, UK:	Acetic acid, Coomassie blue, Glycerol, Glycine, Methanol, Molecular High range marker (205 kDa to 29 kDa), Molecular Low range marker (66 kDa to 14.2 kDa), Ponseau S stain solution, Sodium deocyl sulphate (SDS), Tris-Hydrochloride (pH 6.8, pH 8.8), Alkaline phosphatase linked secondary antibody (Rabbit)
BDH, UK:	Acylamide/Bis-acrylamide (19:1)
NEB, UK:	Protein molecular weight markers
Novagen, UK:	Protein molecular weight markers

Whatman, UK:	3MM Filter paper
Fuji, UK:	Hyperfilm RX-100

Protein expression and purification

Novagen, UK:	pet32a vector,
	Expression host strain :
	BL21 (DE3) [Genotype: F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])].
	BL21 (DE3) [Genotype: F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])], pLysS.
	AD494(DE3) [Genotype: Δ ara-*leu*7697 Δ *lacX74* Δ *phoAPvuII phoR* Δ malF3 F'[*lac*+ (*lacIq*) *pro*] *trxB*::*kan* (DE3).
NEB, UK:	pTwin Impact vector system, Expression host strain ER25666 [(Genotype: F- λ- fhuA2 [lon] ompT lacZ::T7 gene 1 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm]]

Polyclonal-antibodies

Sigma-Aldrich, UK:	Alkaline phosphatase-conjugated secondary Ab, Freund's complete adjuvant, Freund's incomplete adjuvant.
Beckton Dickenson, UK:	1 ml syringes

2.1.3.6 Stock solutions

General:

EDTA: 0.5 M Ethylenediaminetetraacetic acid, pH 8.0

TBE (10X): 1.0 M Tris, 0.9 M Boric Acid, .01 M EDTA(pH 8.0)

SSC (20X): 3 M NaCl, 0.3 M Sodium Citrate (pH 7.0)

TE (1X): 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0)

Loading Dye: 0.25% Bromophenol blue (w/v), 30% Glycerol, TE (1X)

Sequencing Dye: 0.16% (v/v) Deionised formamide, 0.84% (v/v) Blue dextran solution (50mg/ml)

Plasmid DNA preparation:

Luria Bertani (LB) broth (litre) 10 g Tryptone, 5 g Yeast extract, 10 g Sodium chloride, 1 N NaOH (pH 7.0),

LB agar: LB broth plus 1.5% (w/v) BBL grade A agar

TES: 50 mM Tris-HCl (pH 8.0), 25% Sucrose, 2 mM EDTA (pH 8.0)

M-STET: 5% Triton X-100, 50 mM EDTA (pH 8.0), 50 mM Tris-HCl (pH 8.0), 5% TES

Jet Quick (Genomed): Plasmid Mini-prep kit.

RNA:

DEPC-H_2O: 0.01% DEPC was added to deionized H_20, mixed thoroughly and allowed to stand overnight in a fume hood at room temperature. The treated water was then autoclaved for 20 min at 15 psi and temperature 120°C to remove any traces of DEPC.

Lysis Buffer (DEPC): 0.1 M Tris HCl, 50 mM EDTA, 50 mM NaCl, 2% SDS, 0.75% β mercaptoethanaol.

20 x NBC: 1 M Boric acid, 20 mM Sodium citrate, 100 mM NaOH (pH7.5).

10 x Loading Buffer: 50% Glycerol, 100 mM EDTA, 0.25% bromophenol blue.

RNA denaturing buffer: 7.5% Formamide, 2.2% Formaldehyde, 0.75% 20xNBC.

Sodium phosphate buffer: 1M (pH 6.8): 13.14 g Na_2HPO_4, 14.82 g NaH_2PO_4 in 200 ml dH_2O.

Protein:

Running Buffer (Per Litre): 15.1 g Tris base, 94 g Glycine, 10% SDS made up to 1 litre with dH_2O.

Sample Buffer (2X): 50 mM Tris-HCl (pH 6.8), 2% β-mercaptoethanol, 40% Glycerol, 0.1% bromophenol blue, 2 % SDS.

Stain Solution: 0.1% Coomassie blue (G250), 40% Methanol, 20% Acetic acid

Destain Solution: 10% Acetic acid

2.2 Molecular Biology Methods

2.2.1 Electrophoresis methods

Agarose gels

Agarose gels were made by boiling 0.8-1.5% agarose (w/v) in 0.5X TBE. The solution was allowed to cool to 60°C and ethidium bromide stain was added to a final concentration of 100 ng/ml. DNA samples were mixed with loading dye (See Materials section). The electrophoresis buffer was 0.5X TBE (45 mM Tris-Borate, 1mM EDTA) and the stained gels were visualised on a UV transilluminator (UVP Inc.).

Formaldehyde-agarose gels comprised 1% (w/v) agarose dissolved in DEPC treated 1X NBC. The solution was cooled to 60°C and formaldehyde added to a final concentration of 1%. The electrophoresis buffer was 1X NBC. Gels were electrophoresed at 60 volts for a period of 4 hr.

Polyacrylamide gels

Denaturing polyacrylamide gels for protein analysis were made using 40% Acrylogel solution mix (19:1). Sequencing gels for the ABI 377A sequencer were prepared by mixing reagents i.e. Diluent and Concentrate (acrylamide-urea), supplied by EC National Diagnostics (US) as recommended by the manufacturer. The volume was made up to 60 ml with H_2O and polymerised by addition of 400 µl 10% APS and 50 µl TEMED. Sequencing gels were allowed to polymerise for 2 hr. at room temperature.

2.2.2 Transformation of *E. coli* competent cells

E. coli DH5-α competent cells were prepared by the method of Mandel and Higa (1970). Initially a single colony of DH5-α was inoculated in 10 ml LB broth and was aerated for 12 to 18 hr or overnight at 37°C. Then 1 ml of the

culture was added to 100 ml LB in a conical flask and shaken at 220 rpm until O.D. 600 had an absorbance of 0.3 to 0.5. Cultures were subdivided into 10 ml volumes and centrifuged at 4,000 rpm for 10 min in a Sorvall RC5B centrifuge at 4°C. The supernatant was removed and the pellet was resuspended in 10 ml of ice-cold 0.1 M $CaCl_2$ and allowed to incubate on ice for 15 min. The cooled cells were centrifuged at 4,000 rpm for 10 min and the pellet was resuspended in 1 ml 0.1 M $CaCl_2$ containing 15-20% glycerol and stored at -70°C.

Regarding transformation, 40-70 ng of vector DNA from a ligation reaction (ca. 10 µl) was added to 100 µl competent cells and incubated on ice for 15 min. The reaction mix was heat-shocked at 42°C for 1 min followed by returning the samples to ice for 1 min. LB broth (350 µl) was added and the mixture incubated at 37°C for 1 hr. The mixture was centrifuged at 10,000 rpm for 10 min and all of supernatant was removed except for the final 100µl which was used to resuspend the pelleted cells. The mixture was plated on LB agar containing the relevant antibiotic (ampicillin 100 µg/ml or kanamycin 50 µg/ml).

2.2.3 Plasmid DNA isolation for cloning and sequencing

Plasmid DNA was prepared for sequencing and/or cloning using either the Qiagen plasmid minipurification kit according to the manufacturer's protocol or a modified version of the original boiling lysis plasmid method (Holmes and Quigley, 1981). A single colony of *E. coli* was used to inoculate 10 ml of LB containing the relevant antibiotic (ampicillin 100 µg/ml or kanamycin 50 µg/ml) and was incubated at 37°C and shaken at 220 rpm for 12 to 18 hr. The cells were harvested by centrifuging at 3,000 x g at 4°C for 10 min. The supernatant was drained off and the pellet was resuspended in 60 µl TES buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 25% sucrose) and transferred to a 1.5 ml eppendorf tube. A 20 µl volume of fresh lysozyme (Sigma) (20 mg/ml in 250 mM Tris-HCl (pH 8.0)) was added to the resuspended pellet which was incubated on ice for 10 min. 500 µl M-STET lysis solution (5% sucrose, 50 mM Tris-HCl pH 8.0, 50 mM EDTA, 5% Triton X-100) was added to the tube and the sample was allowed to stand at room temperature for 15 min. The tube was placed in a boiling water bath for 1 – 1.5 min and was then centrifuged at 13,000 rpm for 20 min. The glutinous pellet was removed with a sterile toothpick. The supernatant was treated with 10 µl heat-treated RNAase (10 mg/ml) at 37°C for 15 min. The

supernatant was then phenol extracted and the upper aqueous phase was transferred to a new eppendorf tube. Plasmid DNA was precipitated by the addition of 0.6 volumes isopropanol and incubated at room temperature for 15 min. The DNA pellet was collected by centrifugation at 13,000 rpm for 10 min. The pellet was washed with 70% ethanol and air dried. The dried pellet was resuspended in 30 µl sterile water.

Spin Dialysis

Spin Dialysis of samples was used to de-salt the sample, to remove phenol or other small molecules from the DNA solution. Glass beads and Sepharose CL6B were sterilised and equilibrated in 1X TE. A small hole was punctured in a plastic 0.5 ml eppendorf, 60 µl glass beads was added to the punctured tube followed by 0.5 ml Sepharose CL6B solution. The tube was centrifuged in a swing out bucket centrifuge for 2 minutes at 2,000 rpm. The column was then washed with 60 µl sterile water and re-centrifuged for 2 minutes at 2,000 rpm. The column was placed in a 1.5 ml eppendorf which was used as a collection tube, a 50 µl DNA sample was added to the column, and the column was centrifuged at 2,000 rpm for 2 min. The recovered de-salted DNA solution was used in other procedures without further purification.

2.2.4 Sequencing

Sequencing of plasmids or PCR products was performed using the Applied Biosystem (ABI) PCR Cycle Sequence Dye Terminator kit. The cycle parameters for the sequencing reactions were: 96°C for 30 sec, 50°C for 15 sec, 60°C for 4 min - for 25 cycles. Sequencing gel Quartz plates were cleaned with hot Alconox detergent, allowed to dry and wiped with a lint-free tissue soaked with 90% isopropanol. The following EC National Diagnostics (US) reagents were used to make 6.0% polyacrylamide-urea sequencing gels: 39.6 ml Diluent, 14.4 ml Acrylamide-urea concentrate, 6 ml running buffer, 400 µl 10% ammonium persulphate (APS) and 25 µl TEMED. The gel mix was carefully dispensed between the sequencing plates using a 50 ml syringe and allowed to polymerise for approximately 2 hr. Gels were checked by laser scanning for low background fluorescence prior to use. Precipitated PCR sequencing reactions were resuspended in 3 to 4 µl formamide blue-dextran dye, denatured at 95°C for

2 min, snap cooled on ice, loaded onto the sequencing gel and electrophoresed for a period of 16 - 18 hr at 1500 volts. The output sequence trace and text files were formatted and analysed using the ABI373A sequencing software.

2.2.5 Southern blotting

Radiolabelling

DNA fragments and plasmids were radiolabelled with [α-^{32}P] dCTP by the random priming method of Feinberg and Vogelstein (1983) or using the Rediprime labelling kit II (Amersham) under the conditions recommended by the manufacturer.

For method of Feinberg and Vogelstein (1983), 13 µl DNA (at 3 ng/µl) was denatured in a boiling water bath for 5 min and snap cooled on ice. An equal volume of reaction mix (400 mM Hepes pH 7.0, 40 µM each dGTP, dATP, dTTP, 10 U/ml random hexamer primer, 40 mM β-mercaptoethanol, 100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$) was added to the denatured DNA sample followed by 1µl BSA (10 mg/ml) and 1 - 3µl [α-^{32}P] dCTP. The reaction was incubated at room temperature for a period of 4 hr to overnight. The radiolabelled probe was recovered free of unincorporated label by spin dialysis on a 1X TE-equilibrated Sepharose CL6B mini column (Murphy and Kavanagh, 1988).

Hybridisation

DNA restriction fragments were transferred from 0.8% agarose gels onto Nylon hybridization membranes by blotting as described by Southern (1975). The gel was soaked in Denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 15 min and was then neutralised by shaking for 30 min in a Neutralisation solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5). The capillary method was used to transfer DNA fragments onto the membrane with the use of 20X SSC (3M NaCl, 0.3 M tri-sodium citrate) and the membrane briefly washed in 2X SSC. DNA was then fixed onto the membrane by UV crosslinking for 3-5 min.

Pre-hybridisation of DNA fixed to the membrane was conducted at 65°C in 20 ml Hybridisation buffer (6X SSC, 10X Denhardt's solution (2 g/L Ficoll, 2 g/L PVP, 2 g/L BSA), 0.5% SDS, 100 µg/ml heat-denatured Salmom sperm DNA) contained in a Hybaid roller bottle for a period of 4 hr. The purified radiolabelled probe was denatured at 95°C for 5 min, placed on ice for 2-5 min

and then added to 10 ml of pre-warmed hybridisation buffer. The pre-hybridisation solution was then replaced by the probe mix and hybridisation was carried out at 65˚C overnight.

Washing of the blot

The membrane was washed at 65°C in 2X SSC, 0.1% SDS for 15 min followed by 1X SSC, 0.1% SDS for 15 min. and a final wash at 0.5X SSC, 0.1% SDS for 15 min. The membrane was then sealed in Saran wrap and exposed to X-ray film at -70°C.

Stripping of blots

Probes were stripped from membranes by washing in 0.1X SSC, 0.1% SDS at 95°C for 3 min. The membranes were then washed 2 – 4 times at room temperature.

2.2.6 Northern blotting

RNA preparation and capillary blotting

Total RNA samples of 10 – 20 µg (1 µg/µl) were denatured by heating to 65˚C for 15 min with an equal volume of RNA sample buffer (4% formamide, 3.7% formaldehyde, 1X NBC, 12.5 µg/ml ethidium bromide). Samples were then placed on ice for 2 min, centrifuged briefly, stored on ice and 3 µl RNA loading buffer was added prior to loading on an RNA denaturing gel (section 2.1). The RNA was then transferred to a Hybond N+ nylon membrane by capillary blotting with 20X SSC. RNA was fixed to the membrane by exposing to UV on a transilluminator for 5 min.

Hybridisation

Radiolabelled DNA probes were prepared as described in Section 2.2.5. Probe hybridisations were carried out as follows: the blot was placed in a Hybaid rolling bottle and prehybridised in 10 ml pre-warmed hybridisation solution (0.1 g BSA Fraction V, 4 ml dH_2O, 0.7 g SDS, 5 ml 1M sodium phosphate buffer (pH 6.8) and 20 µl 0.5M EDTA, (pH 8.0) for 60 min at 65°C. The denatured

radiolabelled probe was chilled on ice and added to fresh pre-warmed hybridisation solution. Hybridisation was carried out overnight at 65°C.

Washing of blot

The blot was given two separate washes with Wash A solution (0.5% BSA Fraction V, 5% SDS, 40 mM sodium phosphate buffer (pH 6.8), 1 mM EDTA) at 65°C for 10 min each, followed by a further 2 to 4 washes with Wash B solution (1% SDS, 40 mM sodium phosphate buffer (pH 6.8), 1 mM EDTA) at 65°C, each for 10 min. After washing the blot was placed into a cassette with X-ray film and intensifying screens and exposed at –70°C until a strong signal was detected.

2.2.7 Reverse transcription (RT)-PCR

Total RNA (1-4 µg/µl) pre-treated with DNAase-I was added to 100-200 pmol random primers (Hexamer $d(N)_6$), denatured at 70°C for 10 min and placed on ice for 2 min. To the RNA-primer mix was added 1X first strand buffer (25 mM Tris-HCl pH8.3, 37.5 mM KCl, 1.5 mM $MgCl_2$), 10 mM DTT, 0.625 mM dNTPs, 40 units RNAsin (Promega). The reaction mixture was incubated at 20-25°C for 10 min and preheated to 42°C for 2 min before adding 200 units of Superscript II reverse transcriptase (Invitrogen). The cDNA synthesis reaction was carried out at 42°C for 50-60 min and the reaction terminated by heating to 70°C for 15 min. The cDNA (1 µl) was used directly for PCR and produced product sizes up to 1.5 kb. PCR reactions contained dNTPs each at 0.25 mM, oligo primers at 100 ng/µl, 2.0 mM Mg^{2+}, 1X Reaction buffer, 2.5 units Platinum *Taq* DNA polymerase (Invitrogen). Amplification was for 30-35 cycles of denaturation at 94°C for 1 min, annealing at 60°C for 1 min, DNA synthesis at 72°C for 1 min, followed by a final extension at 72°C for 10min.

2.2.8 Total cellular DNA extraction

Total DNA was extracted from *L. perenne* using the method described by Kiang *et al.*, (1993). One to two grams of young leaf tissue was homogenised in 1 to 2 ml of Extraction buffer (100 mM NaCl, 100 mM Tris-HCl (pH 8.0), 0.5 M EDTA, 70 mM beta-mercaptoethanol, 2% SDS) by means of a pestle and mortar. To the resulting slurry was added 1-2 ml phenol (saturated with 50 mM Tris (pH 8.0)) and the mixture was homogenised vigorously for 60 sec. The slurry was

dispensed into eppendorf tubes and centrifuged at 13,000 rpm for 5 min in a microfuge. The top aqueous phase was recovered, placed into a fresh tube and was then extracted two more times with phenol. The final upper aqueous phase was recovered and 2/3 volume of cold Isopropanol was added to precipitate the nucleic acids at 4°C. The nucleic acids were pelleted by centrifuging at 13,000 rpm for 10 min and the pellet was washed with 80% ethanol and air-dried. The pellet was redissolved in 30 µl of sterile water.

DNA samples in 100 mM Tris-HCl pH 7.5 were treated with DNAase-free RNAase (9 µg) in a final volume of 270 µl at 37°C for 20 min. This was followed by two phenol extractions and precipitation of DNA by adding a 1/10 volume of 3M sodium acetate (pH5.2) and 2 volumes of 100% ethanol and storing samples at -20°C for 1 hr. DNA precipitates were recovered by centrifugation at 13,000 rpm for 10 min, the pellet was washed with 80% ethanol and air-dried. DNA was finally resuspended in 30 µl of sterile water.

2.2.9 Total RNA extraction

Total RNA was extracted from *L. perenne* and *N. tabacum* by the method of Eggermont *et al.*, (1996). Approximately 1 to 2 g plant material was homogenised to a fine powder in liquid N_2 using a pestle and mortar. This was quickly transferred to a second mortar containing 1-2 ml of DEPC-treated Lysis Buffer (100 mM Tris-HCl (pH 8.0), 50 mM EDTA, 50 mM NaCl, 2% SDS, 0.75% beta-mercaptoethanol) and 2 ml phenol. This mixture was quickly homogenised and then carefully transferred to eppendorf tubes and centrifuged at 13,000 rpm for 10 min. The upper aqueous phase underwent a single phenol extraction followed by two more phenol:chloroform extractions and a final chloroform extraction. The upper aqueous phase was recovered and was mixed with 0.5 volumes DEPC-treated 8M LiCl and incubated at –20°C in order to precipitate total RNA. RNA was pelleted by centrifugation at 13,000 rpm for 10 min at 4°C. The pellet was washed with 70% ethanol and resuspended in 30 µl DEPC-treated sterile water and stored at –70°C.

2.3.10 Protein analysis

2.2.10.1 Electrophoresis

Denaturing SDS-polyacrylamide gels comprised a separating gel accounting for the bottom 2/3 of the gel and an upper 5% stacking gel. The separating gel was made by mixing 7.5 ml 40% Acrylogel (19:1) with 7.5 ml 1.5 M Tris-HCl (pH 8.8), 0.4 ml 10% SDS, 0.4 ml 10% APS, 12 µl TEMED, and adjusting the total volume to 30 ml with sterile deionised water. The upper stacking gel was made of 1.25 ml 40% Acrylogel (19:1), 1.25 ml 1.0 M Tris-HCl (pH6.8), 6.75 ml H_2O, 0.1ml 10% APS, 0.1ml SDS, 10 µl TEMED. Protein samples were mixed with 2X SDS sample buffer (2% SDS, 2% beta-mercaptoethanol, 50% glycerol, 5 mM EDTA, 0.1% bromophenol blue) and denatured at 95-100°C for 2 minutes before loading. Protein gels were electrophoresed in 1X SDS Running buffer.

2.2.10.2 Protein expression in *E. coli*

Two different protein expression vector system were used: the pTWIN IMPACT II system (NEB) and the pET system (Novagen). The gene-of-interest was cloned into pET32a (Novagen) or the pTwin I vector (NEB). The resulting pET32a clones were transformed into the expression strain BL21(DE3) whereas pTwin I clones were transformed into the host strain ER2566 (a derivative of BL21). In either system optimal conditions of protein expression were investigated by using a variety of parameters including different culture growth temperatures (37°C, 30°C, 25°C, 18°C, 16°C) or concentrations of the inducer: 0.2 mM – 1 mM IPTG. Depending on the solubility of the protein, the thioredoxin reductase negative host strain AD494(DE3) was used with a view to enhancing the solubility of the protein.

2.2.10.3 Purification of an insoluble protein by electro-elution

Insoluble recombinant proteins were fractionated by electrophoresis in a denaturing 10% SDS-polyacrylamide gel and the protein band of interest was excised from the gel using a sterile blade. A series of 4 x 1 cm gel slices were placed into a Biorad 422 electro-elution glass chamber which contained a nominal 10 kDa cut-off membrane at the base. The glass chamber was filled with elution buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS) and was placed into an elution

buffer tank. A volume of 100 ml elution buffer was placed in the upper gel tank and 600 ml of the elution buffer was placed in the lower gel tank. The glass chamber/gel slice was allowed to elute at 10 mA for a period of 4 to 6 hr only. To remove SDS from the elution buffer in the glass chamber, the SDS elution buffer in the upper and lower tank was replaced midway through the run with fresh elution buffer minus SDS. A further step to replace the elution buffer minus SDS with PBS (pH 7.1) was undertaken with the use of nominal 10 kDa retention spin columns.

2.2.10.4 Generating polyclonal antibodies in rabbits

Prior to injection with recombinant protein, a pre-bleed of 10 ml was obtained from the rabbit (day 0). A volume of 1 ml recombinant protein (antigen) was mixed with 1 ml Freund's complete primary adjuvant (final concentration 300 µg/ml protein) in a glass bijou vessel by vigorously shaking for 2 to 4 hr at 4°C which resulted in an emulsion. After this period the homogenous solution was carefully drawn into a 1 ml plastic syringe and was capped with a hypodermic needle. The protein:adjuvant mixture was injected into rabbits by the BioResources Unit (BRU), TCD. After 14 days a 1st booster injection was prepared by vigorously homogenising the recombinant protein in incomplete adjuvant at a final concentration of 250-350 µg/ ml. After 21 days the 1st bleed (10 ml) was collected (BRU, TCD). A 2nd booster injection was prepared after 28 days. Bleeds from rabbits were continuously collected over a six-week period. In all cases the bleeds were left to coagulate for 1-2 hr at 37°C, the blood serum was carefully recovered, and dispensed into fresh tubes. Sodium azide was added to each sample as a preservative to a final 2 mg/ml. The serum was clarified by passing through a sterile 0.45 micron filter and was then stored at –20°C.

2.2.10.5 Western blotting

SDS-PAGE and western blotting was carried out as described in Sambrook *et al.* (1989). Protein samples were resolved on denaturing 10 – 14% SDS-PAGs. The gels were stained in Coomassie blue stain solution (0.1% Coomassie blue (G250), 40% Methanol, 20% Acetic acid) for 30 min by gently shaking at room temperature. Gels were destained in 10% hot acetic acid for 30 min at 60°C by shaking gently and replacing with fresh destaining solution every

10 min. A duplicate SDS-PAG was not stained in Coomassie blue but rather proteins were transferred onto a nylon membrane using Transfer buffer (0.02 M Tris pH 8.0, 0.15 M glycine, 0.02% SDS, 20% methanol) in a wet transfer apparatus according to the manufacturer's instructions (Invitrogen). The efficiency of transfer was determined by staining the membrane with Ponceau S solution and then destaining with sterile dH_2O. The blot was incubated in blocking solution (10 mM Tris-HCl (pH 8.0), 350 mM NaCl, 0.5% Tween-20, 2-9% Non-fat Milk) for 1 hr at room temperature or overnight at 4°C. The blot was washed twice with Solution A (10 mM Tris-HCl (pH7.5), 150 mM NaCl, 0.5% Tween-20, 2-9% Non-fat Milk), primary antibody was added, diluted in 5 ml of Solution A blocking buffer (ranging from 1:10 to a 1:1000 dilution) and incubated with the blot for 1-2 hr with gentle agitation. The blot was washed 3 to 4 times with blocking buffer A and then incubated with the secondary antibody (alkaline phosphatase (AP)-conjugated anti-rabbit (Sigma)) at 1:1000 dilution in 5 ml Solution A. The blot was incubated by gently shaking for 1-2 hr at room temperature. The blot was washed 3 x 5 min in Solution A, and was then incubated in AP Buffer (100 mM Tris-HCl (pH9.5), 100 mM NaCl, 5 mM $MgCl_2$ with gentle agitation for 30 min at room temperature. The membrane was drained of APB. The reagents NBT and BCIP (66 µl each) were added to 10 ml fresh AP buffer, which was added to the blot and allowed to incubate at room temperature for colour development. The blot was then washed in dH_2O, was air dried, and was wrapped in cling film and stored in the dark at room temperature.

2.2.10.6 Affinity purification of antibodies

Recombinant protein (1 mg) was fractionated on a preparative SDS-PAG, transferred onto a nylon membrane by standard western blotting procedures, and the transferred protein visualized by Ponceau S staining. The target band was then cut from the western blot as a strip of membrane, destained in sterile water and allowed to air dry. The protein strip was incubated with shaking in 20 ml 100 mM Glycine-HCl (pH 2.5) for 5 min. The membrane was then washed in TBS buffer (20 mM Tris (pH 7.4), 500 mM NaCl, 0.05% Tween-20). The membrane strip was blocked with 3% Non-Fat Milk in 20 ml TBS by gentle shaking for 1 hour at room temperature. A 1/5 dilution of anti-serum in 10 ml TBS was allowed to bind to the membrane for 2 hr at room temperature in a 14 ml conical Falcon tube with

constant shaking. The supernatant was removed and the membrane was washed with TBS followed by PBS (20 mM sodium phosphate (pH 7.2), 150 mM NaCl). To elute the affinity-purified antibody, the membrane was incubated in 100 mM Glycine-HCl (pH2.5) for 10-15 min with gentle shaking at room temperature. The antibody-containing solution was adjusted to pH 7.0 using 1 M Tris-HCl (pH 8.0). Sodium azide was added to a final concentration of 2 mg/ml and aliquots were then stored at -20°C. The purification and specificity of the affinity-purified antibody was assessed by western blotting.

CHAPTER 3

Analysis of a Novel CMS Gene, *orfC9*, in the Mitochondrial Genome of *Lolium perenne*

3.1 Introduction

To date, 13 mtDNA rearrangements giving rise to chimeric genes have been associated with CMS in diverse species (including *L. perenne*), as illustrated in Figure 3.1 (adapted from Hanson and Bentolila, 2004). Remarkably, in most cases, the chimeric gene is either located immediately adjacent to an ATP synthase (*atp*) gene or contains *atp* coding sequences.

3.1.1 CMS chimeric genes associated with ATP synthase loci

Comparative RFLP analysis of CMS lines of sunflower containing the PET1 cytoplasm (*Helianthus annuus*) with lines containing the fertile cytoplasm (*H. petiolaris* and *H. annuus*), revealed a 17 kb mtDNA rearrangement comprising a 5 kb insertion and a 12 kb inversion which in turn was flanked by 261 bp inverted repeats (Horn and Friedt, 1999; Schnable and Wise, 1998). The 5 kb insertion contained an *atpA* gene and downstream from this, a chimeric orf of 522 bp (*orfH522*) most likely generated by a recombination event (Kohler *et al.*, 1991; Schnable and Wise, 1998). The *orfH522* transcript (Moneger *et al.*, 1994) and the encoded 16 kDa protein were shown to accumulate in a tissue-specific manner in the CMS line.

In maize, the CMS-T-associated gene *T-urf13* was identified by different approaches (Levings, 1993). One involved a comparison of *in organello* translation products synthesized by purified mitochondria from CMS-T and fertile lines. This revealed the synthesis of an abundant 13 kDa polypeptide, named T-URF13 (for T-unidentified reading frame 13) specifically in mitochondria from CMS-T plants. In order to identify the mtDNA region encoding T-URF13, a cosmid clone bank of CMS-T mtDNA was differentially screened with labelled mRNA from CMS-T and fertile mitochondria. Restriction mapping, northern blotting and sequencing of the candidate T-specific region

revealed a pair of co-transcribed orfs *T-urf13* and the conserved mitochondrial gene *orf221* (Dewey *et al.*, 1986). *T-urf13* is located upstream of *orf221* and is primarily composed of non-coding sequences derived from 26S rRNA fused to *atp6* 5'-sequences which provide an upstream promoter region and the *atp6* translational start codon (ATG), whereas *orf221* encodes a membrane bound protein (Prioli *et al.*, 1993) later identified as ATP4 (Heazlewood *et al.*, 2003).

An important approach in confirming the role of *T-urf13* in CMS-T sterility, apart from investigations of the effects of fertility restorer genes, was the investigation of tissue culture-induced mtDNA deletion mutations that resulted in reversion to (as distinct from restoration of) fertility in regenerated plants containing the T-cytoplasm. RFLP analysis of 20 fertile regenerants revealed that the T-*urf13* gene was lost by deletion in 19, due to recombination (Levings 1993; Ward 1995; Schnable and Wise, 1998). In the single remaining fertile mutant that still contained the T-*urf13* region, the orf was found to encode a truncated T-URF13 protein due to a 5 bp insertion in the coding region (Wise *et al.*, 1987; Schnable and Wise, 1998). Furthermore, the co-transcribed *orf221* reading frame was unaltered in all 20 deletion mutants.

A chimeric CMS gene was generated in *Petunia* following construction of somatic hybrids between fertile and CMS lines using the process of protoplast fusion (Hanson, 1991; Boeshore *et al.*, 1985). The hybrids contained recombinant mitochondrial genomes that subsequently segregated during cell divisions. RFLP analysis of recombinant mtDNA in male-sterile and fertile progeny revealed the presence of a CMS-associated locus containing the chimeric *pcf* (*petunia CMS-associated fused*) gene. This chimeric *orf* consists of portions of the coding regions of *atp9* (at the 5' end), the *coxII* exon and an unidentified sequence termed *urfS* (Young and Hanson, 1987). The 1.2 kb *pcf* region is co-transcribed with the mitochondrial genes *nad3* and *rps12* and in the presence of the nuclear restorer gene *Rf*, the *pcf*-derived transcripts are altered (Young and Hanson, 1987) and accumulation of the encoded 19.5 kDa PCF protein is reduced (Nivison *et al.*, 1994).

Physical mapping studies in *B. napus* have revealed differences in the mtDNA of the male-sterile *pol* cytoplasm and male-fertile *cam* cytoplasm that involves a rearrangement in close proximity to the *atp6* gene (L'Homme and Brown, 1993). The rearranged region contains a CMS gene, *orf224,* specific to the

pol cytoplasm (Singh and Brown, 1991). The accumulation of *orf224/atp6* transcripts is altered by the restorer genes *Rfp1* or *Rfp2* (Jean *et al.*, 1997). Also in *B. napus*, the male-sterile *nap* cytoplasm contains a complex CMS-associated region *orf222/nad5c/orf139,* with the predicted protein from *orf222* having 79% amino acid identity with the sequence of the *pol* cytoplasm ORF224 C-terminal region (L'Homme *et al.*, 1997). Transcripts from the *orf222/nad5c/orf139* region are reduced in abundance in fertility-restored verus sterile lines, thus confirming the involvement of this region in the CMS phenotype.

The mt genome of maize plants carrying the CMS-S cytoplasm contains a repeated region 'R' that contains two chimeric *orfs* (*orf355-orf77*) that are closely associated with the upstream *coxI* gene and S1 linear plasmid-like sequences (Zabala *et al.*, 1997; Figure 3.1)). The pollen *orf77* transcript is incompletely edited at numerous sites resulting in the introduction of a premature stop codon, thereby potentially resulting in the synthesis of a 17 residue peptide, as well as the original 77 residue-ORF77 protein (Gallagher *et al.*, 2002). The chimeric *orf77* contains portions of the *atp9* gene (Figure 3.1), and the putative ORF17 peptide is 71% identical with the C-terminus of ATP9, thus making it a possible competitive inhibitor of the ATP9 protein particularly during pollen development, thereby possibly playing a causal role in CMS-S (Gallagher *et al.*, 2002).

3.1.2 Cytochrome oxidase genes associated with CMS chimeric orfs

In wheat, an interspecific cross between *T. timopheevi* and *T. aestivum* (male parent) yielded a CMS line the mitochondrial genome of which contains a chimeric gene, *orf256*. This *orf*, which is co-transcribed with the downstream *coxI* gene, encodes a chimeric protein containing, at its N-terminus, part of the COX1 protein. Polyclonal anti-ORF256 antibodies detected a novel 7 kDa protein that localized to the inner mitochondrial membrane in CMS lines, that was absent in parental lines and in lines restored to fertility (Song and Hedgcoth, 1994; Hedgcoth *et al.*, 2002).

In rice, the CMS-Boro II recombinant mitochondrial genome contains two copies of *atp6,* compared with normal fertile rice (*Japonica*) which contains only one (Kadowaki *et al.*, 1990; Iwabuchi *et al.*, 1993). A chimeric orf, *orf79,* identified by RFLP analyses of the mtDNA of fertile and sterile lines, is located downstream of the second *atp6* gene (B-*atp6*). The predicted ORF79 protein

contains at its N-terminus a sequence of 19 amino acids that are identical to those of the rice COXI protein (Notsu *et al.*, 2002). The rice *Rf1* nuclear restorer gene alters the size of the 2.0 kb *atp6-orf79* transcript (Iwabuchi *et al.*, 1993; Akagi *et al.*, 1995).

3.1.3 Restoration of fertility

In most CMS systems, the mitochondrial lesion responsible for male sterility can be suppressed or complemented by a nuclear gene (Chase, 2006) called a *Fertility restorer* (*Fr*) or a *Restorer of Fertility* (*Rf*). Fertility restoration is said to be either gametophytic or sporophytic depending on the timing of the restoration event. In gametophytic systems restoration takes place after meiosis, while in sporophytic systems it takes place before. Thus, if a single dominant gene is involved, 50% of pollen grains will be fertile following gameotophytic restoration, while 100% fertility will be observed following sporophytic restoration. Although there are several exceptions, most notably the *Fr* gene of *Phaseolus vulgaris* (Janska *et al.*, 1998) and the maize *Rf2* gene (Cui *et al.*, 1996), most *Rf* genes act post-transcriptionally to affect RNA processing, RNA editing or transcript polyadenylation.

In the CMS-T maize, the sporophytic *Rf1* restorer increases the level of splicing of *T-urf13* transcripts, resulting in an enhancement of 1.6 kb and 0.6 kb transcripts which correspondingly results in a large reduction in the levels of the T-URF13 protein (Wise *et al.*, 1996). In sunflower, restoration of CMS plants requires a tissue-specific polyadenylation of the CMS-associated *orf522-atpA* transcripts (Gagliardi and Leaver, 1999). In *B. napus* there is enhanced processing of the CMS-associated *orf224/atp6* distronic transcript by the nuclear gene *Rfp* such that two monocistronic *atp6* transcripts increase in abundance, thus restoring fertility (Li *et al.*, 1998; Brown, 1999). In sorghum, the CMS-A3 cytoplasm is associated with a novel gene, *orf107* which contains sequences derived from *atp9* at its 5' end, followed by sequence similarity at the 3' end with *orf79* in rice (Tang *et al.*, 1996; 1999). The editing of *orf107* transcripts in fertile restored lines containing the *Rf3* allele is different to that in CMS lines. Differential RNA editing correlates with fertility restoration, but an earlier finding of altered RNA editing was not reproducible when field-grown sorghum plants were examined (Pring and Van Tang, 2001; Hanson and Bentolila, 2004). It is not known whether

an ORF107 protein accumulates differentially in CMS and restored lines (reviewed in Hanson and Bentolila, 2004).

Several restorer genes from diverse species have been cloned during the past 5 years. These include the *Petunia Rf* gene (Bentolila *et al.*, 2002), the rice *Rf1* gene (Wang *et al.*, 2006), the *B. napus Rfo* gene (Brown *et al.*, 2003), the *Raphanus sativus Rfk1* gene (Koizuka *et al.*, 2003) and the *Sorghum* gene *Rf1* (Klein *et al.*, 2005). Remarkably all of these restorers belong to a gene family that codes for mitochondrially-targeted proteins containing pentatricopeptide repeat (PPR) motifs. PPR genes were first implicated in mitochondrial function by their effects on translation of the mRNA encoding *coxI* in *Saccharomyces cerevisiae* and the translation or stabilisation of *coxI* and *coxIII* transcripts in human mitochondria (Manthey and McEwen, 1995; Xu *et al.*, 2004). Increasing evidence indicates that organelle-targeted PPR proteins function as RNA transcript-specific binding proteins that mediate interactions between particular RNA substrates and processing/ maturation complexes. A role for a PPR protein in RNA editing has recently been established for the chloroplast gene encoding NADH dehydrogenase subunit D in *Arabidopsis* (Kotera *et al.*, 2005).

3.2 Chapter Aims

As discussed in Chapter 1, the CMS mitochondrial genome in *L. perenne* (CMS9B290) was a product of an interspecific cross between *F. pratensis* (female parent) and *L. perenne* (male parent), followed by a series of backcrosses to *L. perenne* (Connolly and Wright-Turner, 1984; Kiang and Kavanagh, 1993). RFLP analysis of the mtDNA of fertile and male-sterile *L. perenne* revealed the presence of a unique region of approximately 9.6 kb in CMS lines (named LpCMSi) that was absent in fertile lines. At its leftmost border with mtDNA, and located immediately downstream of the normal *atp9* gene, LpCMSi contains a chimeric gene, *orfC9,* which consists of the first 6 codons of *atp9* fused with sequences of unknown origin. Due to its fortuitous location, *orfC9* is co-transcribed with the upstream *atp9* gene (Kiang and Kavanagh, 1996). In contrast, in fertile *L. perenne* lines, the *atp9* gene is co-tanscribed with a different

downstream orf, *orf-L*. Thus, the male-sterile phenotype might be a consequence of either the presence of *orfC9* or the lack of *orf-L* in CMS lines.

In this chapter, *orfC9* and *orf-L* copy number, transcription and transcript editing was investigated in male-sterile, fertile and restored *L. perenne* lines. In addition, antibodies recognizing recombinant ORFC9 protein were produced and evaluated.

3.3 Materials and Methods

3.3.1 Materials

3.3.1.1 *Lolium perenne* lines

The following lines were utilised:

Sterile lines:	CMS3-St (CMS9B290), CMS1, CMS2, CMS4
CMS line restored to fertility:	CMS-Rst1
Fertile maintainer line:	LPS21
Fertile wild-type line:	FtW1, FtW2

3.3.1.2 Primers

***Atp9* primers:**

atp9F 5' TTAATAGGTGCCGGAGCTGCTACAATTGC 3'

atp9R 5' GGCCATCATTAGTGCAAACAATGC 3'

***OrfL* primers:**

orfLF 5' GGAATGAATGGTAGCCATCATGATC 3'

orfLR 5' GTGTAGTTTGGTGTAGTCTTTGTGS 3'

***OrfC9* primers:**

orfC9-F 5' CTC**GGATCC**ATGTTAGAAGGAGCGAAACTGGAC 3'

orfC9-R 5' ACGT**AAGCTT**ATTCATGCGTTGTTTCTCCCTTGAG 3'

The RNA editing primers:

atp9 (AF) 5' **CGGGATCC**GTTCAGTATTTACTCGATATG 3'

atp9 (AR) 5' **CCAAGCTT**GAACAAGACAGAGAGATATTTC 3'

orfC9 (OF) 5' **CGGGATCC**GAAACAATTAAGTATTTACTCG 3'

orfC9 (OR) 5' **CCAAGCTT**GCAACAACAACTTTACTTATTC 3'

Spacer (SpF) 5' **CGGGATCC**GCCTTTTTAATCTTATTCGTTTTCTG 3'

Spacer (SpR) 5' **CCAAGCTT**GGAATATTTTCTTTGTCGTTGGTTTG 3'

Where extra bases have been added to the primer for restriction digestion purposes those bases are shown in bold.

3.3.2 Methods

3.3.2.1 PCR amplification of *atp9*, *orfC9* and *orf-L* fragments

Amplification of the PCR fragment probes *atp9* and *orfC9* from pCMS60 and *orf-L* from pLMF56 was performed using *Taq* DNA polymerase. The reaction conditions used were: 30 cycles for 95°C for 1 min., 60°C for 1 min., 72°C for 1 min., and a further extension of 72°C for 3 min.

The amplification of RT-PCR fragments for *atp9*, *orfC9* and the intergenic region from single stranded cDNA was performed with the use of Platinum *Taq* DNA polymerase using the following PCR conditions of 28 cycles at 95°C for 1 min., 60°C for 1 min., 72°C for 1 min., and a further extension of 72°C of 3 min.

An *orfC9* fragment containing the restriction sites *Bam*HI and *Hind*III for cloning into the pET32a expression vector was amplified from the pCMS60 template using *Pfu* DNA polymerase and the primers orfC9F and orfC9R. The reaction conditions used were 30 cycles for 95°C for 1 min., 62°C for 1 min., 72°C for 1 min.

3.3.2.2 Recombinant protein expression

IPTG-mediated induction of ORFC9 expression was conducted in Terrific Broth (TB) medium (1.2% Bactotryptone, 2.4% yeast extract, 0.4% glycerol, 10% sterile potassium phosphate buffer (pH 7.0)) as described in Chapter 2.

3.3.2.3 Sequencing

The pET32aorfC9 construct (and a pUC19orfC9 construct) were sequenced using the M13 forward and M13 reverse primers (Chapter 2 Materials and Methods), thus confirming the correct sequence and the correct reading frame.

For analysis of RNA editing, the PCR products for the genes *atp9*, *orfC9* and the intergenic region were directly sequenced for both strands.

3.3.2.4 Extraction of total protein from leaf tissue

Total protein was extracted from leaf tissue by homogenizing the 1 g of plant material in 1.5 ml of protein extraction buffer (60 mM phosphate buffer (pH 7.0), 5 mM EDTA, 1 mM PMSF, 0.1% Triton X-100, 10 mM beta-mercaptoethanol) in a sterile mortar. The homogenate was then dispensed into

Eppendorf tubes, placed on ice and was centrifuged for 5 min at 12,000 rpm in the cold room (4°C) to remove cell debris. An equal volume of supernatant was added to 2X Sample buffer (2% SDS, 2% beta-mercaptoethanol, 50% glycerol, 0.1% bromophenol blue, 5 mM EDTA). Samples were placed in a boiling water bath for 2 min to denature the proteins prior to loading on a SDS-PAGE.

3.3.2.5 Extraction of enriched mitochondrial protein

The extractions were carried out at 4°C unless otherwise stated. Approximately 800 mg of freshly-grown leaf tissue was homogenised using a mortar and pestle in 2 ml of STEDP extraction buffer (12% sucrose, 50 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH8.0), 20 mM DTT, 2 mM PMSF, 0.1% BSA, 0.2% PVP). The homogenate was centrifuged at 5,000 x g so as to remove chloroplasts and cell debris. The supernatant was centrifuged at 16,000 x g for 15 min and the remaining pellet (enriched in mitochondria) was resuspended in 100 µl of STEDP buffer. Samples were prepared for PAGE as described above.

3.3.2.6 Generation of polyclonal antibodies recognizing an ORFC9 fusion protein

The highly expressed insoluble protein was cleaned by 4 to 5 high stringency washes in wash solution (0.1 M sodium phosphate buffer (pH 8.0), 0.01 M Tris-HCl (pH 8.0), 0.1 M EDTA, 0.5 M NaCl, 5 mM imidazole). After each wash the pellet of insoluble protein was centrifuged at 12,000 x g at 4°C. The washed ORFC9 fusion protein (300 µg) was thoroughly resuspended in 2 ml of incomplete Freund's primary adjuvant in a glass bijou by vigorously shaking for 2 to 4 hr at 4°C. The solution was then carefully withdrawn into a 1 ml plastic syringe for injection into rabbits (BioResource Unit, TCD).

3.4 Results

3.4.1 Analysis of *atp9*-associated transcripts in fertile, sterile and restored *L. perenne* lines

To determine if the presence of *orfC9* or the lack of *orf-L* was causally implicated in the CMS phenotype, a comparative northern blot analysis was carried out on total RNA from fertile, sterile and restored lines, using the relevant gene probes (Figure 3.2). Since *orfC9* and *atp9* share 18 bp of homology at their 5' ends, probes lacking this region were amplified from the pCMS60 plasmid: the *atp9* probe comprised 207 bp whereas the *orfC9* probe comprised 352 bp. Hybridization with *atp9*, detected a 1.4 kb transcript in both the sterile line (CMS-St) and the restored line (CMS-Rst1). Hybridizations using the *orfC9* probe detected the same 1.4 kb transcript in both the sterile and the restored line, but did not hybridize with RNA from the fertile maintainer line LPS21 (Figure 3.2A (i)). In contrast, the *atp9* probe detected a 1.0 kb transcript in the fertile line (LPS21) as illustrated in Figure 3.2A(ii). A third probe, *orf-L* (242 bp), encompassing sequences located downstream of *atp9* that are unique to the fertile line was amplified from pLMF56 (see Chapter 4, Figure 4.11). This probe detected the same 1.0 kb transcript in RNA from the fertile line that had initially been detected by the *atp9* probe, but failed to hybridise to any transcripts in the sterile or the restored lines (Figure 3.2A, panel (iii)).

Thus, the northern hybridization data support the observation by Kiang and Kavanagh (1996) that *orf-L* and *orfC9* are co-transcribed with *atp9*, respectively, in fertile lines, and lines containing the CMS cytoplasm. Importantly, the data also indicate that neither the abundance nor the size of the *atp9/orfC9* co-transcript differs between sterile (CMS-St) and fertile restored (CMS-Rst) lines and that reappearance of an *orf-L* transcript is not associated with restoration to fertility.

3.4.2 Investigation of *atp9* and *orfC9* editing in sterile and restored *L. perenne* lines

RT-PCR and sequencing of the amplified products was used to investigate possible editing of the *atp9/orfC9* co-transcript. First-strand cDNA was generated from DNAase-treated total RNA of leaf and flower tissue of sterile, restored and

fertile lines by reverse transcription. The *atp9* gene was amplified from floral and leaf cDNA templates derived from the restored line (but from leaf cDNA only in the case of the sterile and fertile lines) using the primers AF and AR (Figure 3.3). The *atp9* PCR product (288 bp) included 18 bp upstream and 42 bp downstream from the coding sequence. The various *atp9* products were directly sequenced, and in addition, two *atp9* cDNA inserts that had been cloned into the vector pUC19 (using the restriction sites *Bam*HI and *Hind*III; Materials and Methods).

An *orfC9* PCR fragment (429 bp) with extended upstream and downstream sequences of 24 bp and 16 bp, respectively, was amplified from leaf and flower tissue of the restored line but only leaf tissue of the sterile line using the primers OF and OR (Figure 3.3). These PCR products (Figure 3.3) were also directly sequenced along with two cDNA clones representative of the tissue type examined in each line. The intergenic or spacer region between the *atp9* and *orfC9* coding genes using primers SpF and SpR (Figure 3.3), gave a product size of 379 bp that was amplified from leaf and flower tissue of the restored line, but only leaf tissue of the sterile line. PCR products were directly sequenced, followed by sequencing of two cDNA clones. Sequencing of 379 bp spacer products was identical to its corresponding genomic sequence (Figure 3.4).

The sequence data, however, showed no evidence of RNA editing in, sterile or restored fertile lines, in either leaf or floral tissue. Thus, restoration to fertility is not accompanied by differential editing of the *atp9/orfC9* cotranscript (Figure 3.4).

Although editing of *atp9* transcripts does not appear to occur in *L. perenne,* it is nevertheless of interest to compare it with *atp9* sequences from other monocot species. In species in which *atp9* editing occurs, the nucleotide sequence of the edited cDNA is highly homologous to that of *L. perenne* and any bp differences occur at non-edited sites (Figure 3.5). Remarkably, the ATP9 amino acid sequence is 100% identical in all the species shown in Figure 3.5. Thus all of the nucleotide differences found in *L. perenne* are silent substitutions.

3.4.3 Determination of *atp9*, *orfC9* and *orf-L* copy number in CMS and fertile mtDNA

The mt genome of rice CMS lines carrying the Bo II cytoplasm has two copies of *atp6*: a normal copy, N-*atp6*, and a CMS-associated copy, B-*atp6*,

which is co-transcribed with the chimeric gene *orf79*. To investigate whether *L. perenne* lines possess one or more copies of *atp9*, a Southern blot analysis of total genomic DNA from FtW2, and several different CMS lines was carried out. The *atp9* probe detected single bands of 13.2 kb and 6.0 kb, respectively, in *Bam*HI and *Hind*III-digested mtDNAs of CMS3-St (Figure 3.6). Single bands were also detected in the sterile lines CMS1, CMS2, the restored line CMS-Rst1 but with slight discrepancies in size compared to CMS3-St line. A lower molecular weight *Hind*III band of ca. 5.6 kb hybridized with the probe in the fertile line FtW2 although this may have more to do with an electrophoresis phenomenon than an actual size difference. Alternatively, the size variation in hybridizing bands in each line might indicate different mitotypes as has been observed in different lines of carrot (Robison and Wolyn, 2002; Robison and Wolyn, 2005).

The *orfC9* and *orf-L* probes described above were also used in a Southern blot analysis (Figure 3.7) to investigate orf copy number, and their utility in discriminating between CMS and fertile cytoplasms. With the *orfC9* probe, lines containing the CMS cytoplasm gave single bands of 13.1 kb and 5.9 kb, respectively, for *Bam*HI- and *Hind*III-digested samples; no bands were detected by the *orfC9* probe in fertile lines (FtW1 and LPS21). The same blot was stripped and probed a second time using the *orf-L* probe in order to verify the association of *orf-L* specifically with fertile cytoplasms (FtW1 and LPS21). The probe detected single faint bands of 13.2 kb and 6.0 kb, respectively, in *Bam*HI and *Hind*III-digested mtDNA from fertile lines, and no bands were observed in sterile lines; unfortunately, because the previous probe was not completely stripped from the blot, weak hybridization signals due to *orfC9* are still visible in a few of the CMS lanes.

Thus, *orfC9* and *orf-L*, like *atp9*, exist as single copies in, respectively, the CMS and fertile lines.

3.4.4 Expression of recombinant ORFC9 protein in *E. coli.*

In order to determine if the abundant *orfC9* transcripts expressed in CMS lines were translated, it was decided to express recombinant ORFC9 protein in *E. coli*, use it to elicit the production of polyclonal antibodies in rabbits, and use the antibodies for immunodetection of ORFC9 in *L. perenne*.

The *orfC9* coding sequence (378 bp) was amplified from pCMS60 using the primers orfC9F and orfC9R, digested with *Hind*III and *Bam*HI and cloned into pET32a, as confirmed in Figure 3.8A. The pET32orfC9 construct encodes a His6-tagged Trx-ORFC9 fusion protein of 32.5 kDa that was expressed at high level following induction of *E. coli* BL21-DE3 cultures with 1 mM IPTG, at a temperature of 28°C for a period of 3 to 4 hr. (Figure 3.9A, lanes 5 and 6). Sonication of the induced cells followed by centrifugation of the lysed cell extract indicated that the 32.5 kDa Trx-ORFC9 fusion protein was found exclusively in the pellet and was therefore insoluble (Figure 3.9 B, lane 3).

The host strain *E. coli* AD494(DE3), which possesses a thioredoxin reductase activity that reduces the formation of disulphide bonds was used as an alternative expression strain for pET32aorfC9 in the hope that the recombinant protein might adopt a more soluble conformation. However, the Trx-ORFC9 protein was once again found to be insoluble (data not shown).

Hydrophobicity plots for ATP9 and the ORFC9 protein are shown in Figure 3.10. The ATP9 protein is strongly hydrophobic, especially towards the C-terminus, indicative of its location in the mitochondrial membrane. ORFC9 although more hydrophilic than ATP9 contains several strongly hydrophobic regions, for example, between amino acid residues 15 – 25 and 50 - 65 (Figure 3.10), suggesting that it might function as a membrane spanning protein. Moreover, the high hydrophobicity of the ORFC9 protein might also explain its tendency to form insoluble inclusion bodies when expressed in *E. coli*. A common finding for other CMS-associated chimeric proteins is their hydrophobic nature which reflects their location in the mitochondrial membrane, for example, T-URF13 in maize.

Given the insoluble nature of the Trx-ORFC9 fusion protein, purification was attempted under denaturing conditions in 6–8 M urea. The fusion protein contains a His6 tag enabling purification by nickel ion affinity chromatography. However, the urea-denatured fusion protein did not bind efficiently to the Ni-matrix, as evidenced by the quantity of Trx-ORFC9 protein in the unbound fraction. For this reason very little could be recovered during the specific elution steps (Figure 3.11 A, lane 7).

Because affinity purification of the fusion protein proved to be problematic, it was decided to isolate inclusion bodies by centrifugation and

enriching for the insoluble fusion protein by washing the inclusion body pellet 4 – 5 times. This unpurified, washed fraction was used to generate the anti-Trx-ORFC9 polyclonal antibodies.

3.4.5 Western blot detection of ORFC9

The efficiency with which the polyclonal antibody could detect the Trx-ORFC9 fusion protein was evaluated by western blot analysis of *E. coli* cell extracts using an antibody dilution of 1/1000 (Figure 3.12). The antibody detected as little as 50 ng of the 32.5 kDa Trx-ORFC9 fusion protein (Figure 3.12; lanes 7 to 11). It also detected a band of approximately 34 kDa in extracts of uninduced cells that contained the pET32aorfC9 construct but only when large quantities of protein were loaded (20 to 50 μg per lane) (Figure 3.12; lanes 2 – 4). This most likely represents cross-reactivity with an *E. coli* protein that contaminated the washed inclusion bodies used as the antigen. Finally, as expected, the antibody detected the Trx protein expressed from the pET32a vector lacking the *orfC9* insert (Figure 3.12, lanes 5 - 7).

Total leaf protein was extracted from the CMS-St and CMS-Rst lines and the fertile maintainer line LPS21, fractionated by PAGE, transferred to a nylon membrane and probed with the anti-Trx-orfC9 antibody (1/100 dilution). As shown in Figure 3.13 A, the antibody did not reveal the presence of the predicted 14 kDa ORFC9 protein in either the CMS-St or CMS-Rst lines. Similarly, no bands of the expected size were detected in protein samples enriched for mitochondrial proteins (Figure 3.13 B).

3.5 Discussion

The identification of *orfC9* as a candidate CMS-determining gene in *L. perenne* was initially based on a comparison of its structure and that of chimeric orfs in other CMS systems, and in particular, their close association with *atp* genes. CMS-determining genes in petunia (Pruitt and Hanson, 1991), maize CMS-S (Zabala *et al.*, 1997) and sorghum (Tang *et al.*, 1999) contain sequences derived from *atp9* while in *Brassica* (Handa *et al.*, 1995; Landgren *et al.*, 1996), and sunflower (Kohler *et al.*, 1991; Sabar *et al.*, 2003), chimeric orfs contain sequences from *atp6* or *atp8.* In maize atp6 gene provides 5' regulatory sequences for the chimeric *urf13*-T gene (Dewey *et al.*, 1986). Co-transcription of CMS-determining genes with essential mitochondrial genes is also frequently observed in many of these systems.

Due to the high frequency of recombination in plant mt genomes, chimeric orfs are not uncommon. For this reason, caution must be exercised when considering a candidate chimeric orf as a causal factor in male-sterility. Comparison of the fertile and CMS-Owen mt genomes of *B. vulgaris* revealed four chimeric orfs that are independently transcribed in CMS-Owen and are absent from normal mitochondria. However, none of these orfs have so far been proven to be causal factors in the CMS phenotype (Satoh *et al.*, 2004). Clearly, therefore in the case of *L. perenne*, comparative investigations of *orfC9* and *orf-L* expression in male-sterile, fertile and restored lines were required to shed light on their contribution, if any, to male sterility.

In the fertile maintainer line LPS21, *orf-L* is co-transcribed with *atp9* giving a single 1.0 kb transcript. This raised the question as to whether or not the absence of *orf-L* in CMS lines, corresponding to a loss-of-function mutation, is responsible for the male-sterile phenotype. Our findings suggest that it is not, because *orf-L* transcripts are not observed in the CMS-Rst line (Figure 3.2). Since most cases of CMS involve gain-of-function mutations, a more plausible scenario might be that male sterility resulted from the acquisition of *orfC9*. The mt genome of *F. pratensis*, the original cytoplasmic donor parent, might have carried *orfC9* as a silent or latent gene that was subsequently transferred to the *L. perenne* mt genome by recombination. *OrfC9* expression may then have been activated in its new context i.e. the *L. perenne* nuclear background, resulting in male sterility. In

B. tournefortii, for example, a chimeric gene that was silent in mitochondria of the donor cytoplasm was nevertheless expressed in an alloplasmic situation where it caused male sterility (Landgren *et al.*, 1996). Similarly, *B. napus* lines that contained *Arabidopsis* mitochondria showed deregulation of mt orfs that were silent in their original nuclear background but were expressed in the *B. napus* nuclear background, resulting in male sterility and other developmental abberations (Leino *et al.*, 2003; reviewed in Pelletier and Budar, 2007). Thus altered nucleo-cytoplasmic interactions in alloplasmic lines can result in male sterility.

Neither the size nor the abundance of *atp9/orfC9* transcripts was altered, at least in leaf tissue, following restoration to fertility. This contrasts with several other CMS systems, in which restoration to fertility correlates with (i) an altered transcriptional pattern e.g. in the case of the maize CMS-S *orf355* and *orf77* (Wen and Chase, 1999), rice (BoII) *atp6/orf221* (Iwabuchi *et al.*, 1993) and wheat *orf256* (Song and Hedgcoth, 1994); or (ii) a significant decrease in the accumulation of CMS gene transcripts as observed for maize *T-urf13* (Wise *et al.*, 1999), *B. napus orf224/atp6* (Li *et al.*, 1998), *B. napus* (nap) *orf222* (Brown, 1999), sorghum (A3) *orf107* (Tang *et al.*, 1996).

There are, however, several examples of CMS systems in which neither the accumulation nor the processing of CMS-determining transcripts is altered by *Rf* genes. The CMS-specific *orf138-atp8* co-transcripts of *B. napus* CMS-Ogura (Bellaoui *et al.*, 1999) and the CMS-specific *orf125-atp8* co-transcript of Kosena CMS-radish (Koizuka *et al.*, 2000) are not altered by the *Rfo* gene (Bellaoui *et al.*, 1999) and the *Rfk1* or *Rfk2* restorers, respectively (Koizuka *et al.*, 2000). In both systems the restorer genes acts post-translationally to destabilise the ORF138 Ogura and ORF125 Kosena CMS-associated proteins (Bellaoui *et al.*, 1999; Koizuka *et al.*, 2000, 2003), thus resulting in restored fertility. Furthermore, recent reports in *B. vulgaris* CMS-Owen have also shown that the nuclear restorer genes have no effect on transcript patterns of the four putatively associated CMS determining *orfs* each of which is associated with an essential mitochondrial gene (Satoh *et al.*, 2004).

The effect of Rf genes on CMS-associated transcripts may therefore depend not only on the type of *Rf* gene used but also on whether there is an association between the CMS-determining gene and an essential mitochondrial

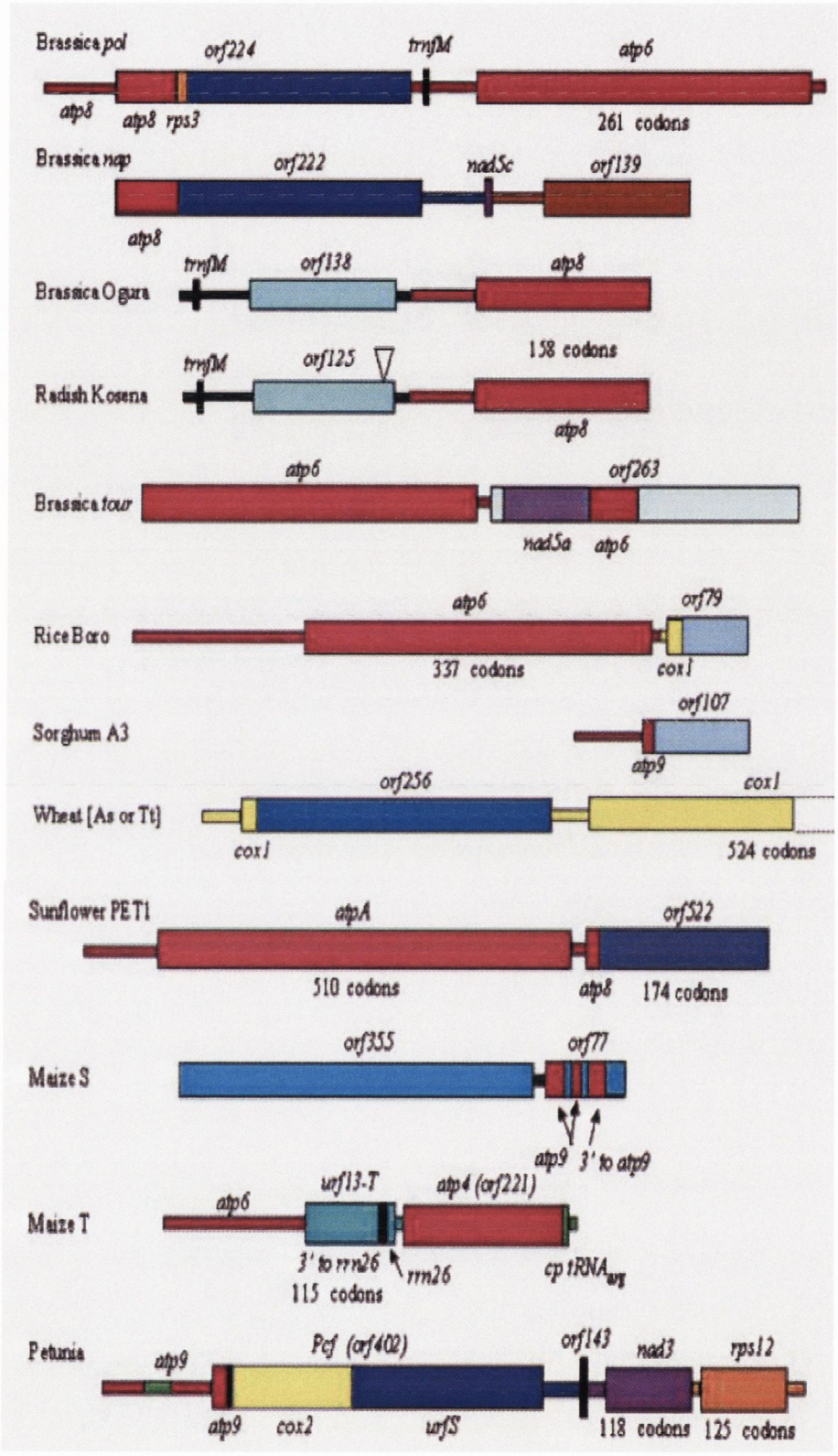
gene. For example, in *B. napus* CMS-Ogura the ATP8 protein can only be translated from the *orf138/atp8* co-transcript (Bonhomme *et al.*, 1992). As hypothesised by Bellaoui *et al.*, (1999), an *Rf* gene that altered the accumulation or processing of this co-transcript might impair the synthesis of ATP8 subunits and have a detrimental effect on plant phenotype. In the case of *L. perenne*, the *orfC9* gene is similarly associated with the essential mt gene *atp9* which exists as a single copy in the mt genome of fertile, sterile and restored lines. Restoration to fertility by a mechanism that altered the processing or abundance of *atp9/orfC9* transcripts might not be tolerated, due to potential negative effects on ATP9 synthesis.

The possibility that differential editing of *atp9/orfC9* co-transcripts might play a role in sterility and restoration to fertility in *L. perenne* was also investigated. However, no differences were detected in a comparison of fertile, CMS-St and CMS-Rst lines. Restoration of fertility by editing of co-transcripts has been intensively investigated in rice carrying the CMS-Boro II cytoplasm. The CMS-Boro II mt genome contains 2 copies of the *atp6* gene, a normal copy, N-*atp6,* and a duplicated CMS-specific copy B-*atp6*, transcribed as a co-transcript with *orf79 (*B-*atp6/orf79).*. In male-sterile plants, the primary B-*atp6/orf79* co-transcript (2.0 kb) is partially edited. However, introduction of the *Rf1* gene results in cleavage of the B-*atp6/orf79* co-transcript to give 1.5 kb and 0.45 kb sub-transcripts (Iwabuchi *et al.*, 1993) and complete editing of B-*atp6* transcript. This prevents the production of cytotoxic peptides and thus restores fertility.

Polyclonal antibodies generated against the Trx-ORFC9 fusion protein failed to detect ORFC9 protein in either total leaf protein extracts or in enriched mitochondrial protein extracts of the CMS-St line and CMS-Rst lines by Western blotting. This does not necessarily imply that ORFC9 protein is not translated in CMS mitochondria or that it plays no role in male sterility. ORFC9 might be expressed at very low levels, below the detection sensitivity of the polyclonal anti-ORFC9 antibodies. Alternatively, ORFC9 protein might be expressed or degraded in a highly tissue-specific manner. It might, for example, be expressed only in tapetal cells or pollen during a narrow developmental window and be absent in leaves in which case an immunohistochemical approach to ORFC9 detection in anthers might be more appropriate. In *Phaseolus vulgaris,* for example, the CMS-associated gene *orf239* is constitutively transcribed in leaf and

reproductive tissue, but the encoded protein is only detected in reproductive tissue, because it is fully degraded in vegetative tissue by proteases (Abad *et al.*, 1995; Sarria *et al.*, 1998). Tissue-specific expression of plant mitochondrial genes has not been widely reported and *orf239* is the only example of a CMS-associated gene regulated post-translationally.

A.



B.

Lolium perenne

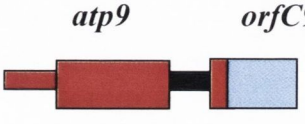


Figure 3.1 The Chimeric genes of the mitochondrial genome associated with CMS in higher plants.

A. The size of the different orfs listed according to the conventional number of codons, except for those based on historical convention as in maize *urf-13* encodes a 13 kDa protein, in petunia *pcf* illustrates CMS associated fused gene, in sunflower *orf522* encodes 522 nucleotides rather than 522 codons. The red colour indicates genes for subunits of the ATP-synthase, shades of blue indicate unknown reading frames within CMS-associated regions whereas shades of yellow indicate genes of subunits of cytochrome oxidase, orange shade indicates ribosomal proteins, brown shade illustrates conserved unidentified reading frames in mitochondria of higher plants and the green shade indicates chloroplast like sequences. This figure is adapted from Hanson and Bentolila (2004). **B**. In *L. perenne* a chimeric gene *orfC9* of 126 codons was identified as being closely associated with ATP synthase gene *atp9*, present in a CMS line CMS9B290 (Kiang and Kavanagh, 1996)

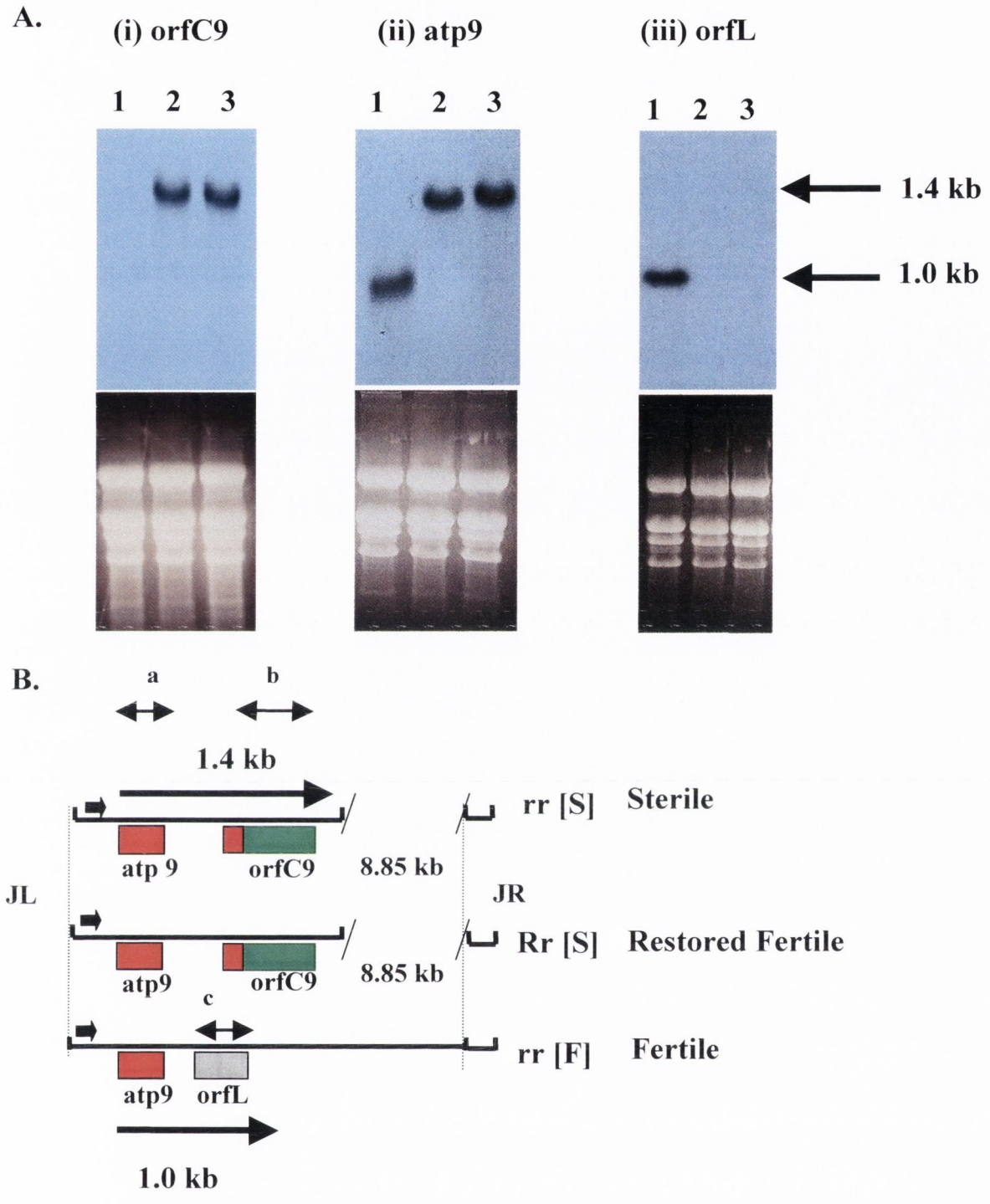


Figure 3.2 Northern blot analysis of mitochondrial transcripts of fertile (LPS21), sterile (CMS9B290), and the restored line (Rst1).

A Northern blot of total RNA was probed with (i) orfC9, (ii) atp9, (iii) orf-L. Lane 1 is fertile-rr[F] RNA (LPS21), lane 2 is restored fertile RNA-Rr[S] (Rst1), lane 3 is sterile RNA –rr[S]. (CMS9B290). **B** The black lines represent the sequence between the junction sites **JL** and **JR.** The location of the open reading frames atp9, orfC9 and orf-L are indicated by rectangular boxes below the lines. The symbol ➡ represents the atp9 promoter and the position of the Northern blot probes **a** (atp9), **b** (orfC9), **c** (orf-L) is indicated by an arrow-bar above the lines. The transcript sizes (kb) are indicated by an arrow symbol.

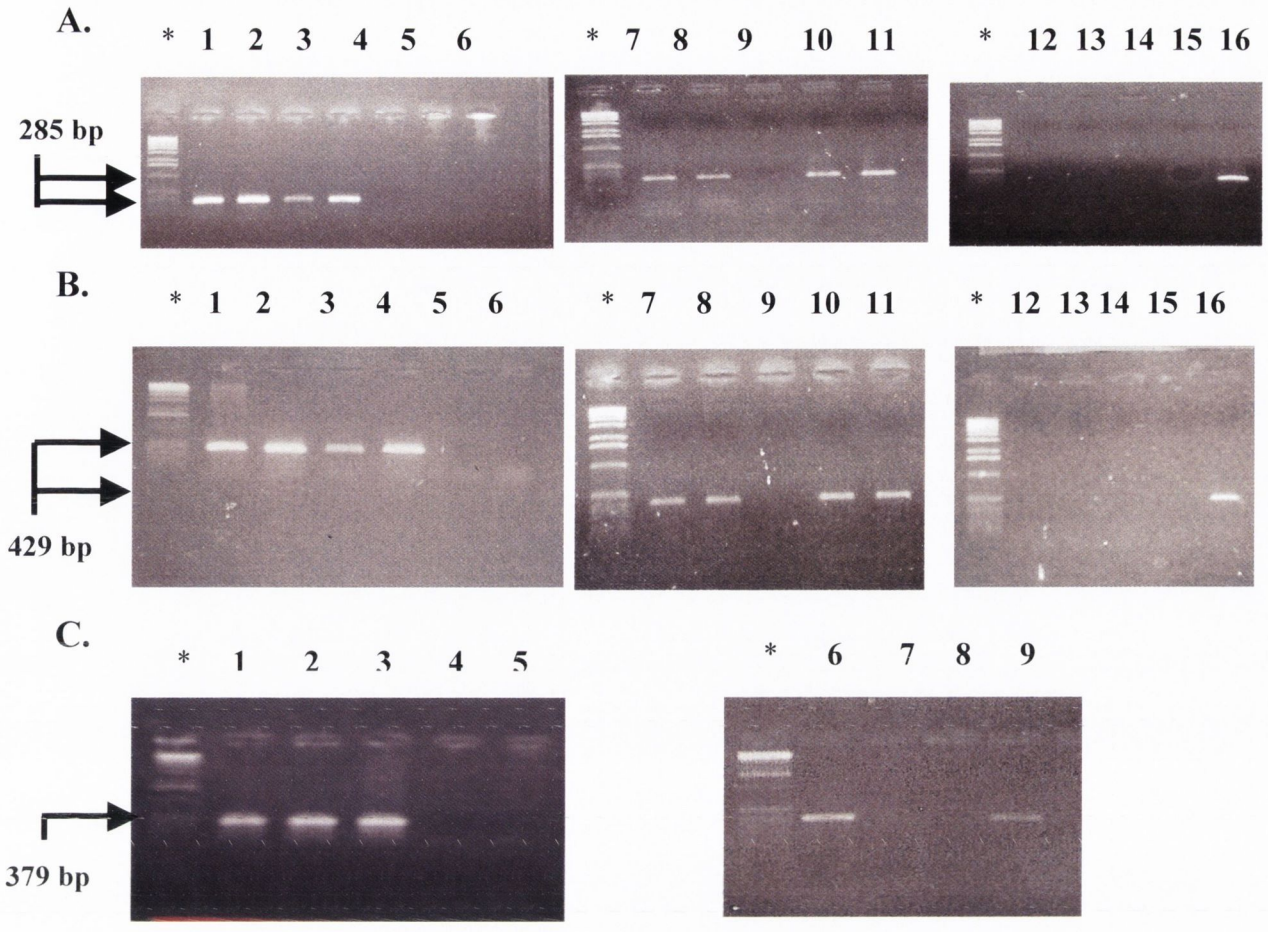
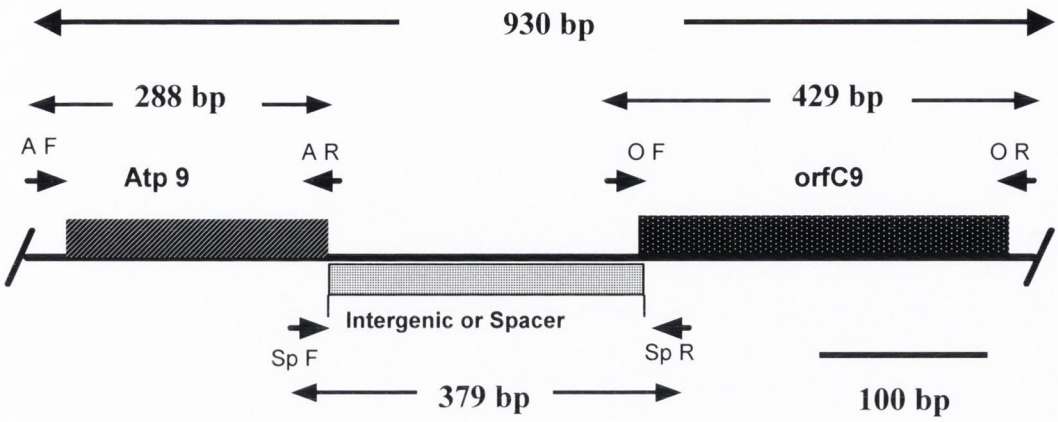


Figure 3.3 PCR products amplified from the cDNAs of *L.perenne* for atp9, orfC9 and the intergenic (spacer) region.

A. Reverse transcribed cDNA (+RT) 285 bp atp9 product in Restored leaf (Rst1) is lanes 1 and 2, Restored fertile flower (Rst1) is lanes 3 and 4, – RT control in leaf and flower is lanes 5 and 6 respectively. Reverse transcribed (+RT) atp9 in fertile wild type leaf (FWt1) lanes 7 and 8, –RT control lanes 12 and 13. Reverse transcribed (+RT) atp9 in Sterile leaf (CMS3-St) lanes_10 and 11, – RT control lanes 14 and 15. Negative control (minus template) is lane 9 and positive control for genomic DNA in sterile leaf (CMS3-St) is lane 16. **B.** Reverse transcribed cDNA (+RT) 429 bp orfC9 product in Restored leaf (Rst1) in lanes 1 and 2, Restored fertile flower (Rst1) in lanes 3 and 4, –RT control in leaf and flower is lanes 5 and 6, respectively. The reverse transcribed (+RT) duplicate of 429 bp orfC9 product is present in Restored fertile leaf lanes 7 and 8, –RT control for leaf is lane 12 and 13. Reverse transcribed (+RT) 429 bp orfC9 product in Sterile (CMS3-St) leaf is lanes 10 and 11, –RT control for leaf is in lanes 14 and 15. The negative control (no template) is lane 9 and the positive control for genomic DNA in sterile leaf (CMS3-St) is lane 16. **C.** Reverse transcribed cDNA (+RT) 379 bp spacer product in leaf and flower tissue for fertile Restored (Rst1) in lanes 1 and 2 respectively, positive control for genomic DNA in lane 3, -RT control lane 4, negative (minus template) control lane 5. Reverse transcribed (+RT) 379 bp spacer product in leaf tissue for the Sterile line (CMS3-St) is in lane 6, -RT control is lane 7, negative control is lane 8, positive control for genomic spacer DNA is lane 9. The 1 kb marker is represented by the lane *.

A.



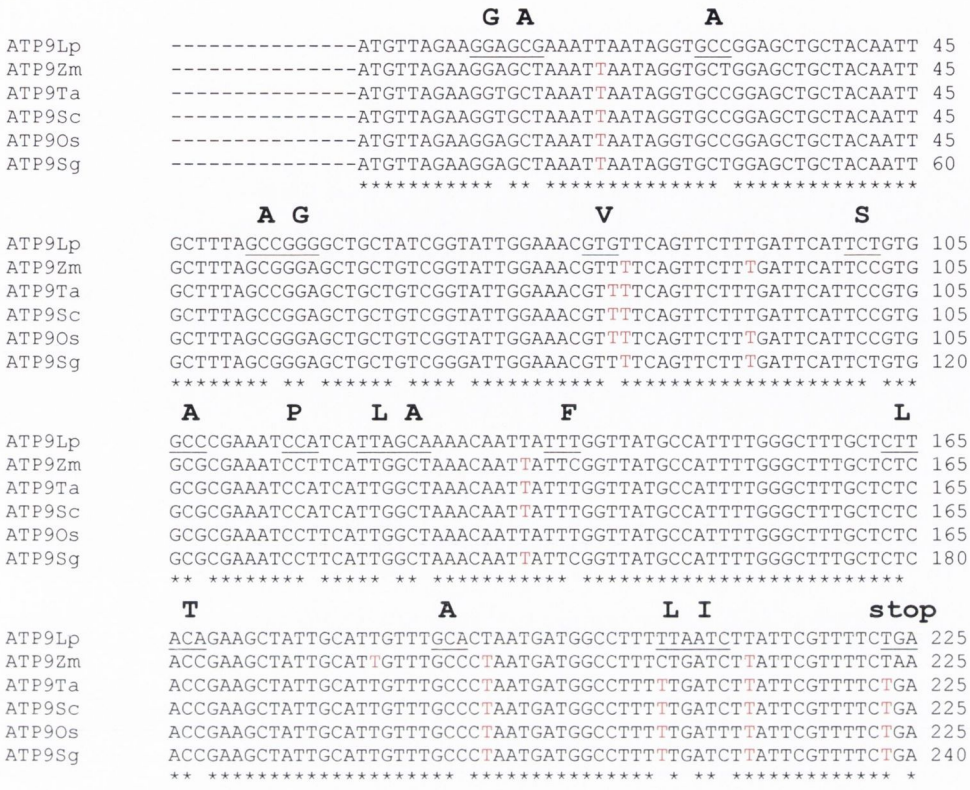
B.

```
      AF                                  atp9
GTTCAGTATTTACTCGATATGTTAGAAGGAGCGAAATTAATAGGTGCCGGAGCTGCTACA 60

ATTGCTTTAGCCGGGGCTGCTATCGGTATTGGAAACGTGTTCAGTTCTTTGATTCATTCT 120

GTGGCCCGAAATCCATCATTAGCAAAACAATTATTTGGTTATGCCATTTTGGGCTTTGCT 180
                                                  SpF
CTTACAGAAGCTATTGCATTGTTTGCACTAATGATGGCCTTTTTAATCTTATTCGTTTTC 240
STOP                                          AR
TGATAGCAGAAAAAGTCAGAAAACGAAATATCTCTCTGTCTTGTTCCGCGGATTGGGCGT 300

CCTATCCGAGGAATGAATGGTAGCCGTCATCATCTTTTCTTTTTCTGGTTATATAATATA 360

TAATATAATACCACGAAAAAATTATTTATATAGTCATTCTCGTGCTAAAGTGGGAGGGGA 420

ATGACAATTCCAAAGAAAAGGTGCTTCTCCTCCGCAGCGGAGTGGAGAAGACGAATCAAC 480
                                                    OF
CAAAAAATACAAGAAGATAGAATGGGAAGAAAAAGCGCGAAACAATTAAGTATTTACTCG 540
      orfC9                                  SpR
ATATGTTAGAAGGAGCGAAACTGGACCCCAAACCAACGACAAAGAAAATATTCCCAGGTT 600

TATCCGGTTTATTCGGGAGCCTGGGCATTATTACCTTCGCCGGGACAACGGCGGGGATAA 660

AGTATTTCTTCGGACTCTCGTCCGTCGAAATGGCTTTACTAGTGGTTATTTCCACTTTTT 720

TCAGTGTTGTTTTCCTCTACCTGACAACAATGCTCTATAAATATTCTCGCTTAATAGAAC 780

TCCTTATTTCAAAGAACATGGAACAGGGTGAGAAACAGACTATCTTTCACGAGTTGGAAT 840

TTTCCCTATATCGTGGGGTCAAACTCACAGTAAAGGGACGAGACCAACTCAAGGGAGAAA 900
          STOP                OR
CAACGCATGAATAAGTAAAGTTGTTGTTGC 930
```

Figure 3.4 PCR products and nucleotide sequence amplified from cDNA in Sterile, Restored Fertile and Fertile lines.

A. RT-PCR analysis of a 930 bp size fragment that includes the atp9, orfC9 and spacer region in a sterile, restored and fertile lines of *L. perenne.* Primer positions are illustrated by the symbol ➔ and cDNA fragment sizes of 285 bp, 379 bp, 429 bp represent atp9, spacer region and orfC9 respectively. **B.** Nucleotide sequence of the 930 bp cDNA fragment and positions of primers (AF and AR), (OF and OR), (SF and SR) used to amplify fragments highlighted by arrows, and the sequence of atp9, orfC9 and the spacer regions is shown. Nucleotide sequence identities between *orfC9* and *atp9* is underlined.

A.



B.



Figure 3.5 Nucleotide sequence alignment of the atp9 cDNA coding sequence in the monocot species *L. perenne* (Lp), *Z. mays* NB (Zm), *T. aestivum* (Ta), *S. cereale* (Sc), *O. sativa* (Os), *S. bicolor* (Sg).

A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bp is noted by the absence of the asterisk symbol *, and the corresponding triplet codon sequence is underlined with the amino acid residue being indicated by a single letter code. TGA or TAA are used as the stop codon. **B.** High conservation of the translated ATP9 cDNA with no alteration in amino acid residue compared with other monocot grass species.

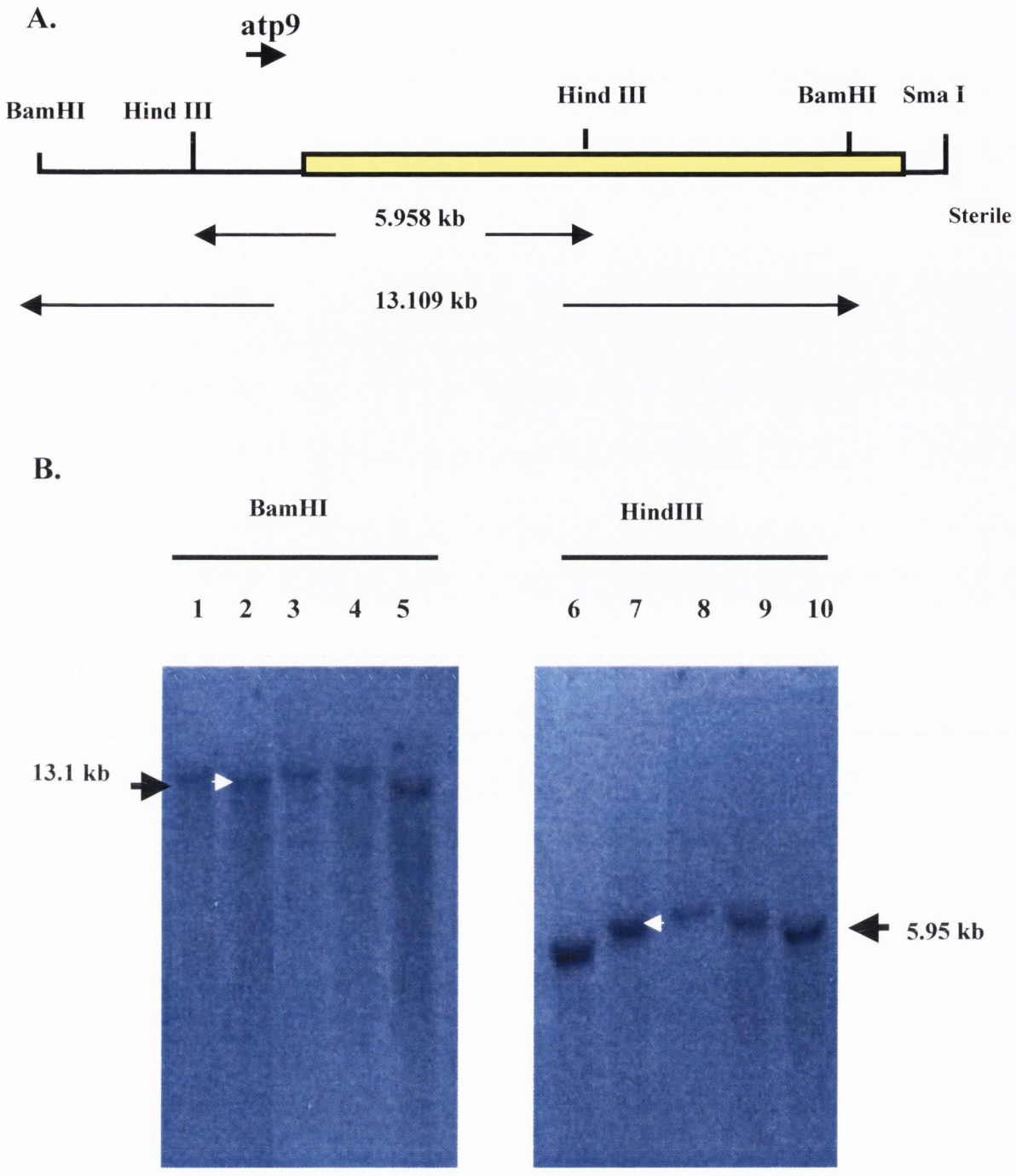
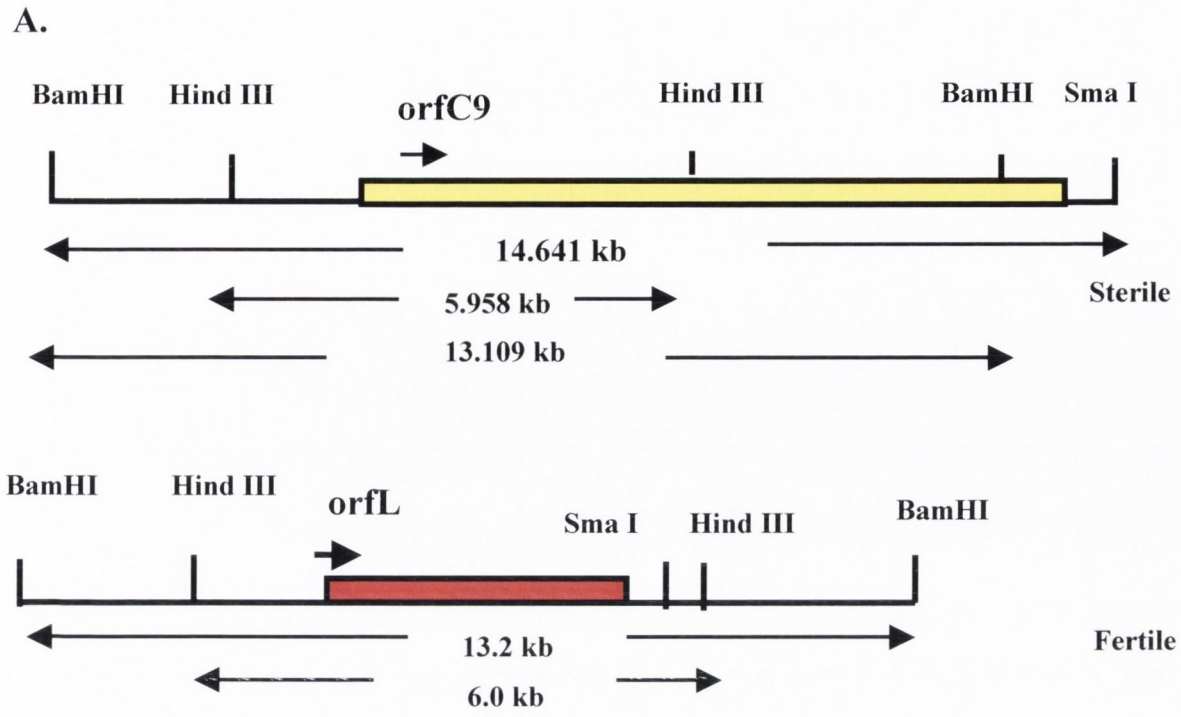
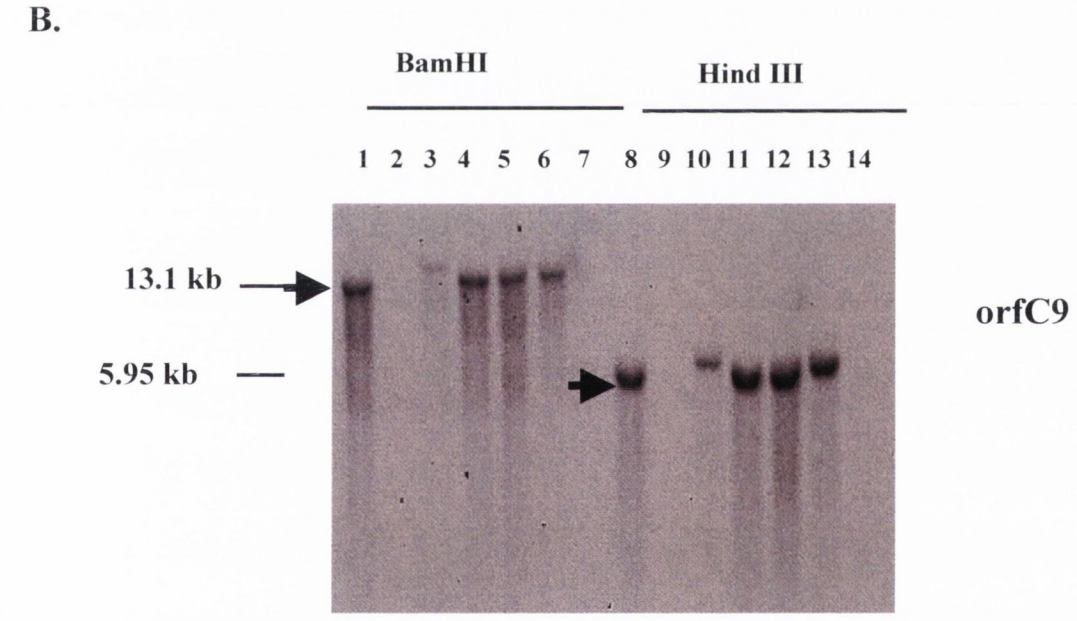


Figure 3.6 Southern blot of *Bam*HI and *Hind*III digested *L. perenne* probed with atp9

A. Diagram of a map showing the size of the *Bam*HI and *Hind*III restriction sites in the Sterile line CMS9B290. The coloured rectangular box represents the non-homologous region specific to the sterile line CMS9B290 cytoplasm. **B.** Total DNA of *L. perenne* digested with *Bam*HI (lane 1 to 5), *Hind*III (lane 6 to 10). Lane 1 Fertile wild type (FtW2), lane 2 CMS3-St (CMS9B290), lane 3 CMS2, lane 4 CMS1, lane 5 Restored fertile Rst1, lane 6 Fertile wild type (FtW2), lane 7 CMS3 (CMS9B290), lane 8 CMS2, lane 9 CMS1, lane 10 Restored fertile Rst1. The 13.1 kb and 5.95 kb size refers to the *Bam*HI and *Hind*III fragment band in lanes 2 and 7 respectively for the sterile line CMS9B290 (CMS3-St).


A.
BamHI
Hind III
orfC9
Hind III
BamHI
Sma I
14.641 kb
5.958 kb
Sterile
13.109 kb
BamHI
Hind III
orfL
Sma I
Hind III
BamHI
13.2 kb
Fertile
6.0 kb


B.
BamHI
Hind III
1 2 3 4 5 6 7 8 9 10 11 12 13 14
13.1 kb
5.95 kb
orfC9

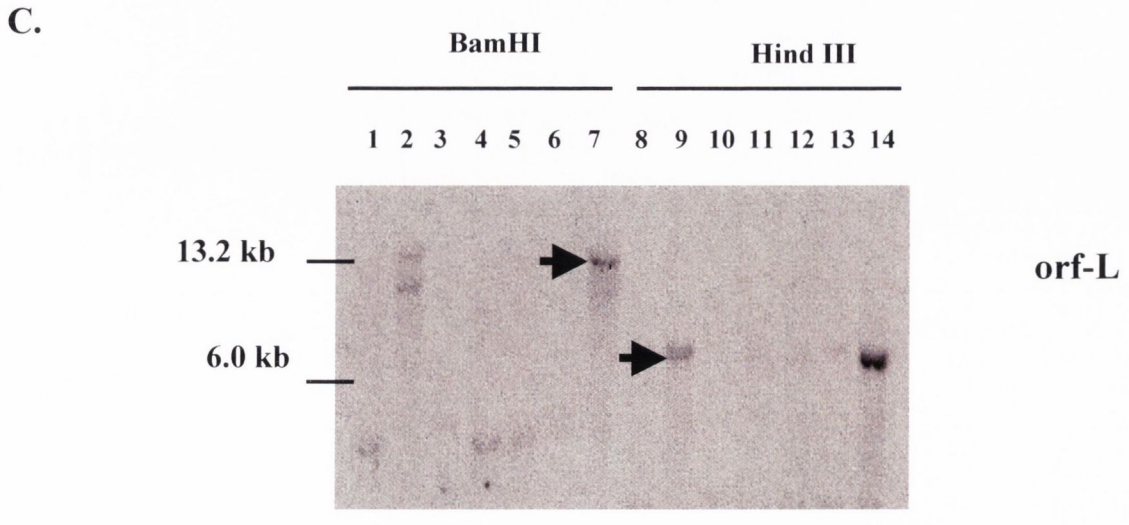

C.
BamHI
Hind III
1 2 3 4 5 6 7 8 9 10 11 12 13 14
13.2 kb
6.0 kb
orf-L

Figure 3.7 Identification of different lines of *L. perenne* containing the CMS cytoplasm by Southern blot analysis.

A. Diagram of a map showing the size of the *Bam*HI and *Hind*III restriction sites. The coloured rectangular box are non-homologous regions and the black lines are homologous regions. Two different size bands 13.1 kb and 13.2 kb in *Bam*HI digested in Sterile (CMS9B290) and Fertile line LPS21 respectively and 5.95 kb and 6.0 kb in *Hind*III digested in Sterile (CMS9B290) and Fertile line (LPS21) respectively. **B**. Total DNA digested with *Bam*HI (lanes 1 to 7), *Hind*III digested (lanes 8 to 14) and probed with orfC9. Lanes 1, 8 restored Rst1, lanes 2, 9 Ftw1, lanes 3,10 sterile CMS1, lanes 4, 11 sterile CMS 2, lanes 5, 12 sterile CMS3-St (CMS9B290), lanes 6, 13 sterile CMS4, lanes 7, 14 fertile maintainer LPS21. **C**. Total DNA digested with *Bam*HI (lanes 1 to 7), *Hind*III digested (lanes 8 to 14) and probed with orf-L; Lanes 1, 8 restored Rst1, lane 2, 9 fertile Ftw1, lanes 3, 10 sterile CMS1, lanes 4, 11 sterile CMS2, lanes 5, 12 sterile CMS3–St (CMS9B290), lanes 6, 13 sterile CMS4, lane 7, 14 fertile maintainer LPS21

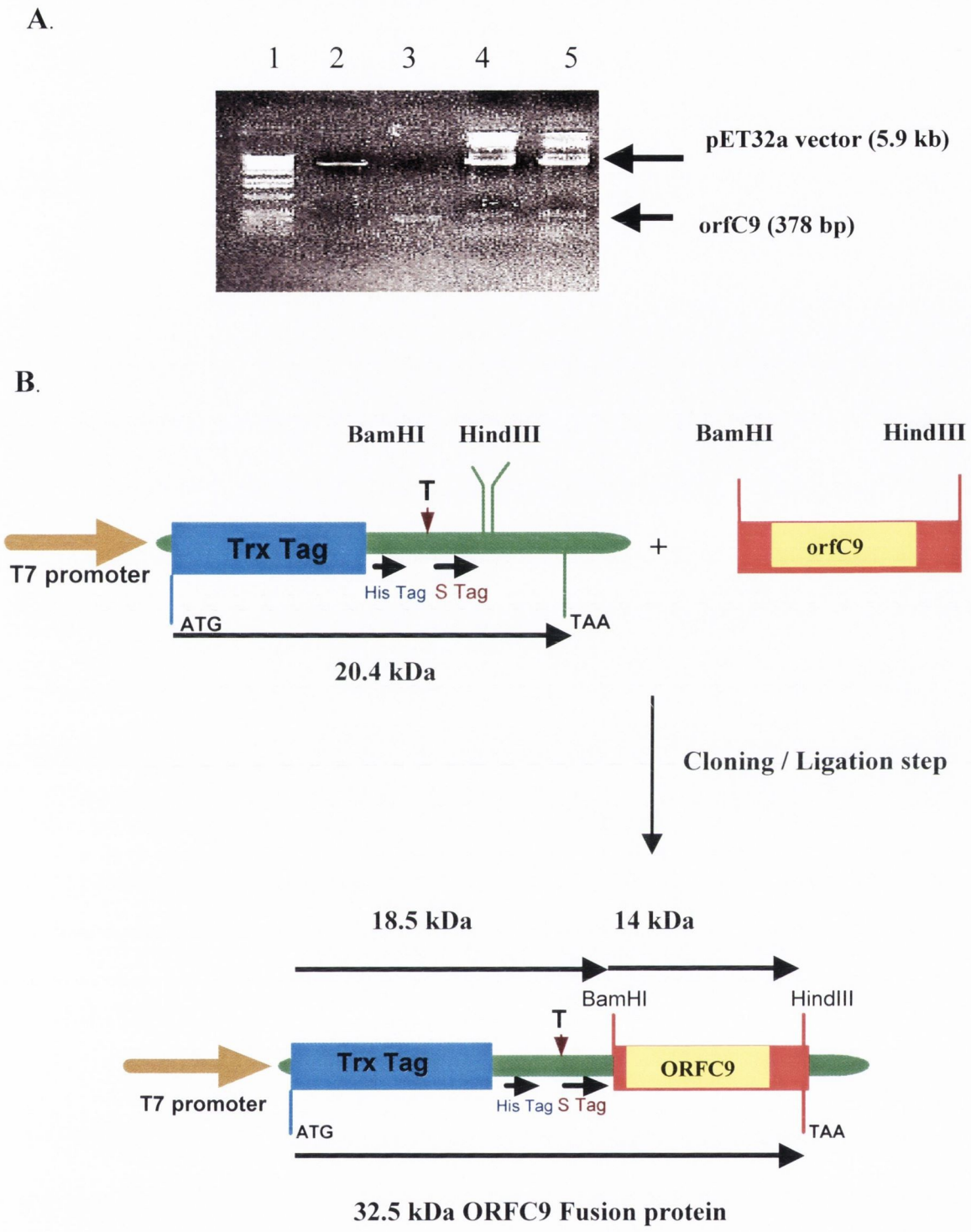


Figure 3.8 Illustrates the insertion of the orfC9 fragment into the pET32a expression vector.

A Restriction digestion of pet32aorfC9 clone. Lane 2 is *Bam*HI/*Hind*III of vector pet32a, lane 3 orfC9 DNA fragment, lanes 4 to 5 are *Bam*HI/*Hind*III digestions of pET32aorfC9 clone, lane 1 is the 1 kb marker. **B** Schema for the cloning of orfC9 into the pet32a expression vector. Promoter used is the T7, the Tags highlighted are the His, S and Trx tags. Thrombin cleavage site highlighted by T, start codon is ATG and stop codon is TAA.

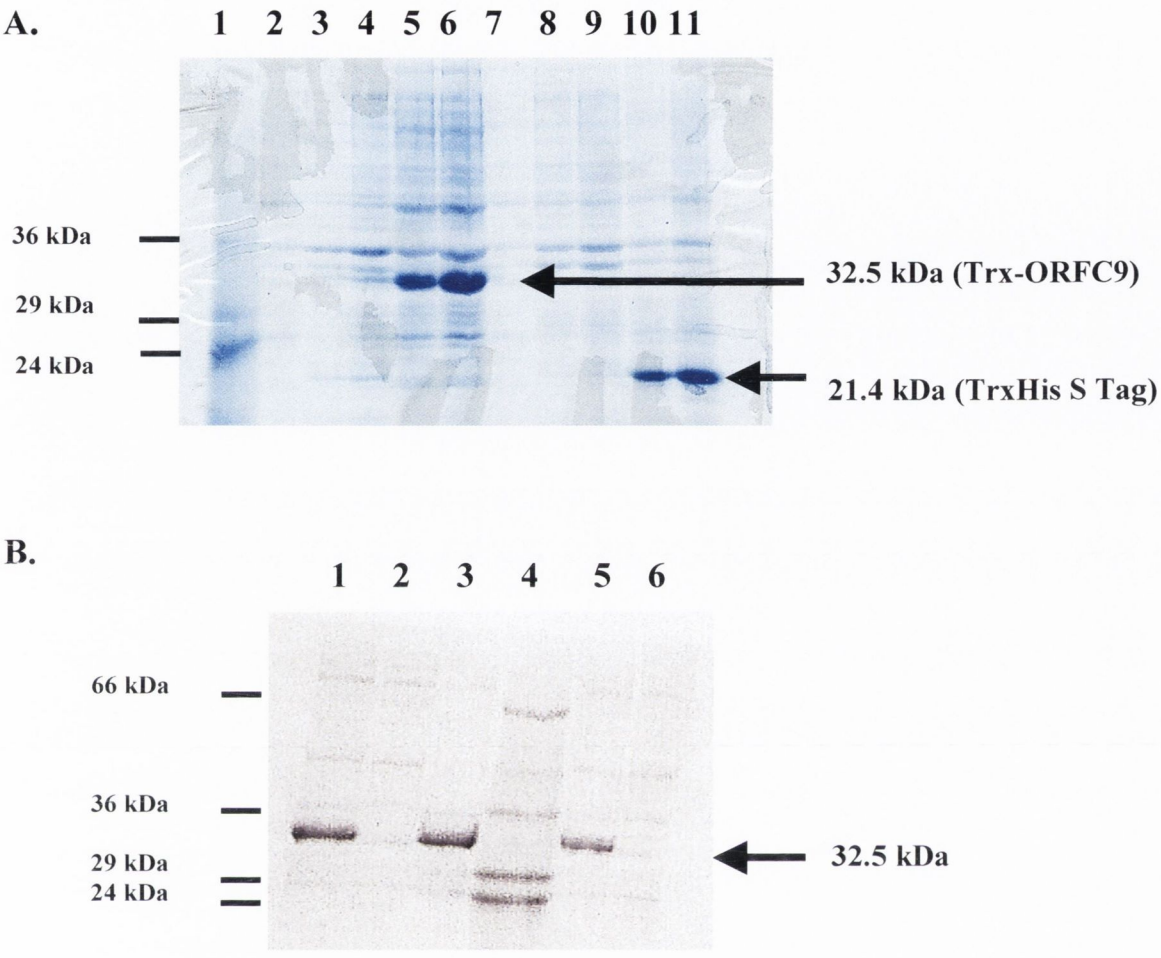
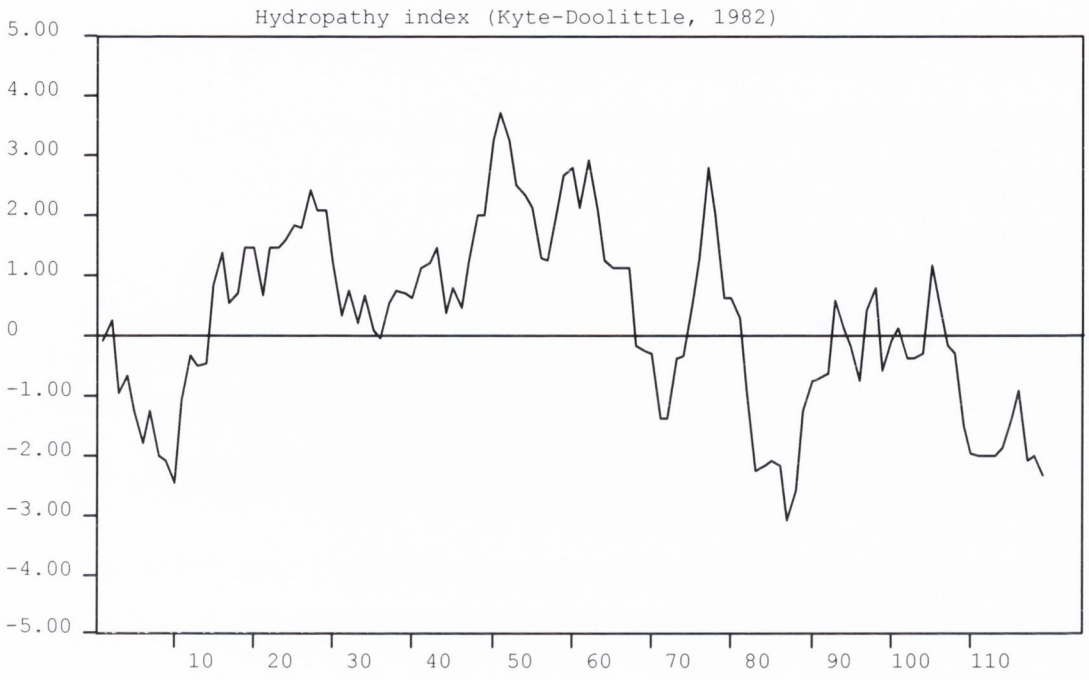


Figure 3.9 Protein expression of ORFC9 protein using the pET32a expression vector system.

A. Is a 10% SDS-PAG stained in Coomassie blue. Lanes 2 to 4 are uninduced (-IPTG) pET32a-orfC9 cells of t=0, t=2hr, t=5hr respectively, lane 5 and 6 are induced (+IPTG) pET32a-orfC9 cells having t=2hr, t=5hr respectively, lanes 7 to 9 are uninduced cells of vector pET32a only of t=0, t=2hr, t=5hr respectively, lanes 10 and11 are induced cells of pET32a with t=2hr and t=5hr respectively. **B.** Solubility of pET32a-orfC9 fusion protein, lane 1 is induced Total cell lysate (**TL**), lane 2 is induced soluble fraction (**SP**), lane 3 is induced Insoluble fraction (**IP**), lane 5 and 6 is induced and uninduced whole cell respectively. Lane 4 Low range Sigma protein marker (kDa).

A.



B.

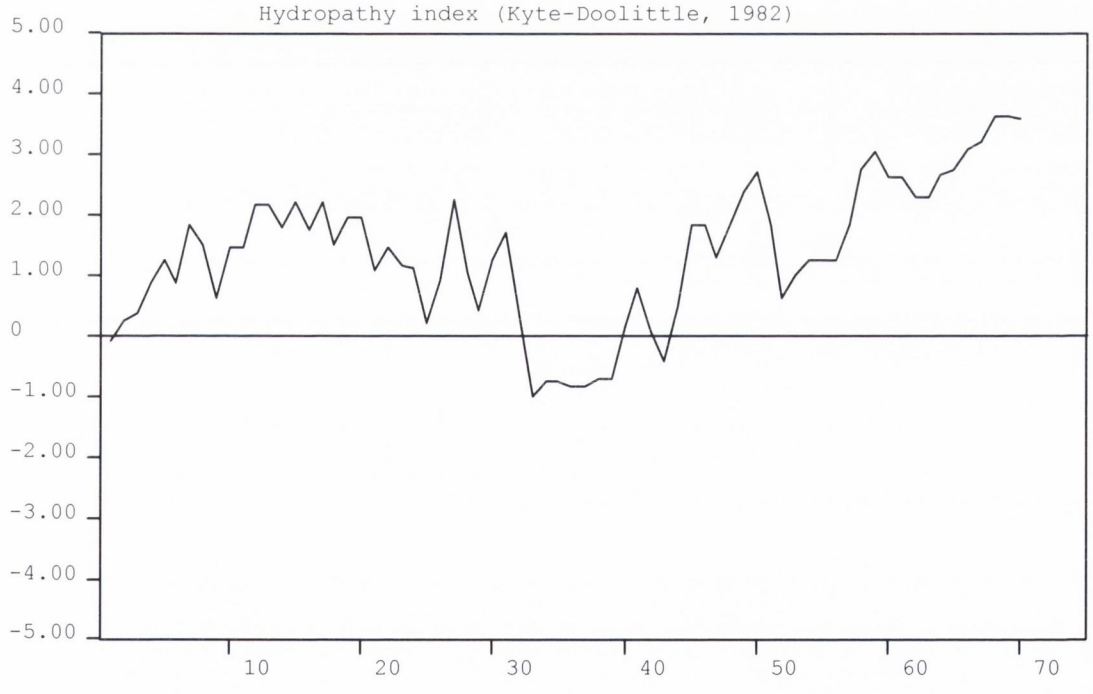


Figure 3.10 Hydrophobicity plot for ORFC9 and ATP9

A. The proposed 14 kDa ORFC9 protein is Hydrophobic with transmembrane potential. **B.** The 8.4 kDa ATP9 protein is highly hydrophobic with location in the inner mitochondrial membrane.

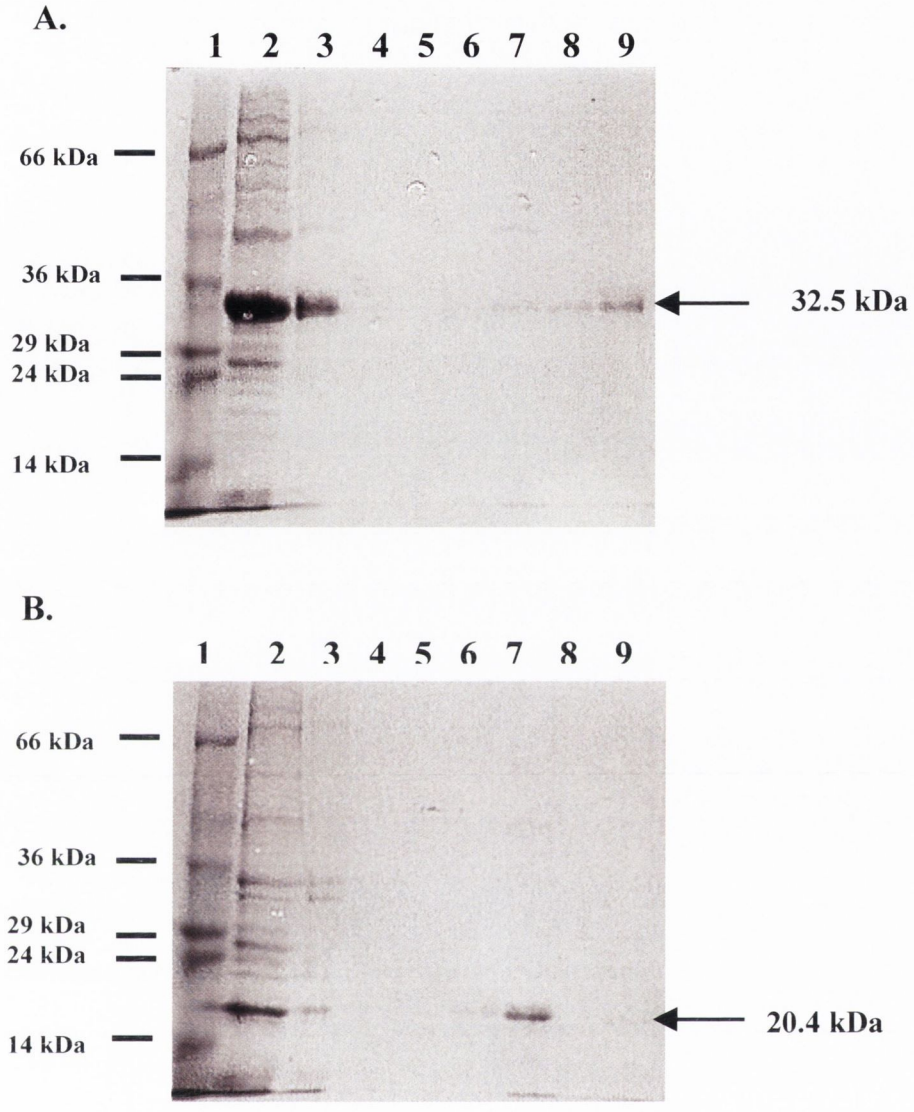


Figure 3.11 Purification of 32.5 kDa Trx-ORFC9 fusion protein and 20.4 kDa TrxHisTag protein under denaturing conditions

A Purification of 32.5 kDa Trx-ORFC9 fusion protein under denaturing (6M Urea) conditions, examined on a 10% SDS-PAG stained in coomassie blue. Lane 2 is insoluble protein, lane 3 is unbound protein, lanes 4, 5, 6, 7 are the elution steps (E4, E3, E2 and E1) and lanes 8 and 9 are the wash steps (W1, W2).

B Purification steps for the 21.4 kDa TrxHisSTag under denaturing conditions, examined on a 10% SDS-PAG stained in coomassie blue. Lane 2 is induced soluble protein, lane 3 is unbound protein, lanes 4, 5, 6 and 7 are the elution steps (E4, E3, E2 and E1) and lanes 8 and 9 are the wash steps (W1, W2). Low range protein molecular marker (Sigma) is present in lane 1.

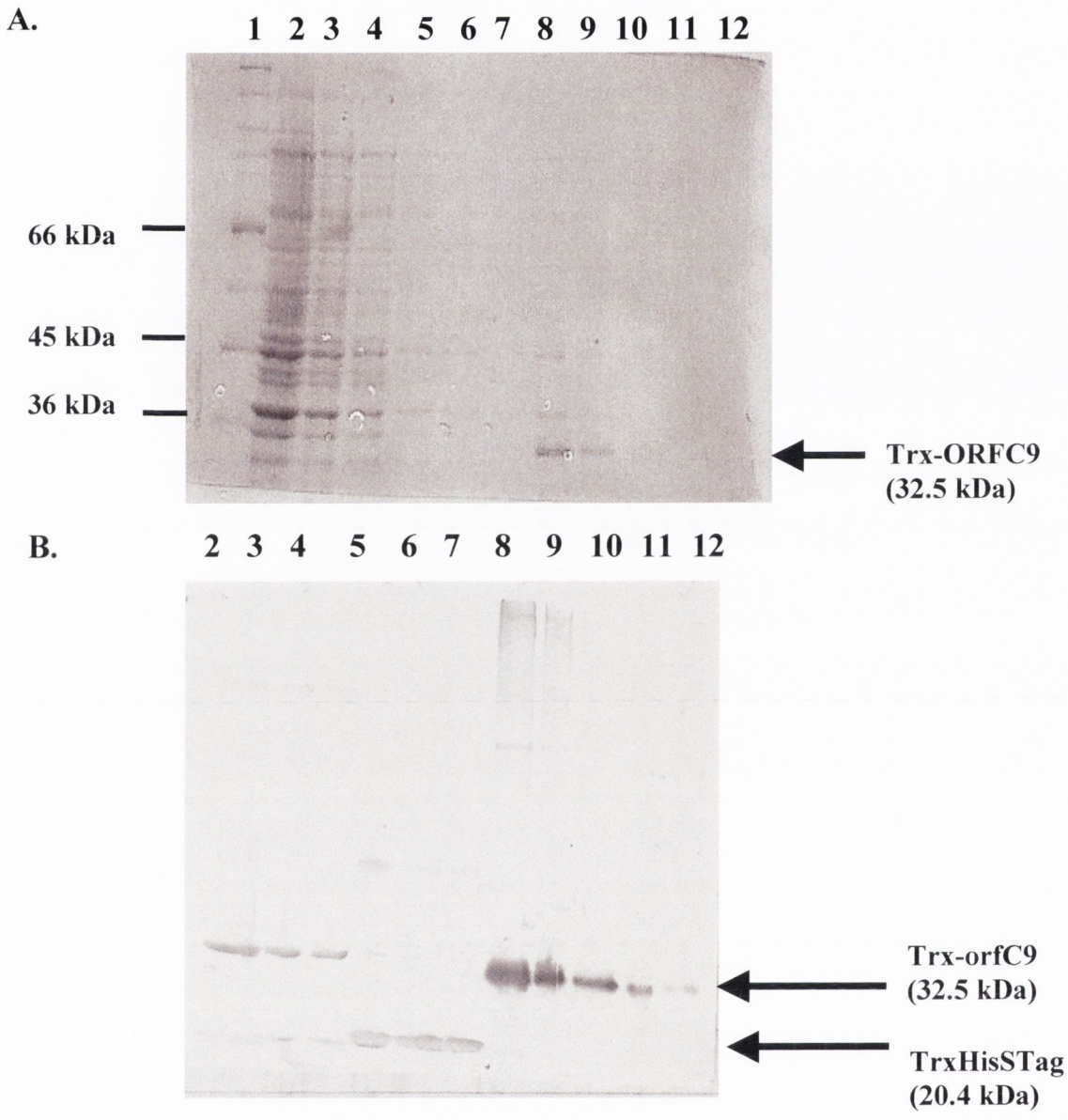


Figure 3.12 Western blot of pET32a recombinant system by probing with the orfC9 antibody

A. 10% Coomassie stained gel of pet32a system, lane 2 (50 µg), lane 3 (30 µg), lane 4 (20 µg) is uninduced pet32aorfC9, lane 5 (5 µg), lane 6 (2 µg), lane 7 (1 µg) is induced pet32a vector only, and lane 8 (4 µg), lane 9 (2 µg), lane 10 (1 µg), lane 11 (300 ng), lane 12 (50 ng) is induced pet32aorfC9. **B.** Western blot of the coomassie stained gel in Figure 3.12A. The electroblot is probed with anti-Trx-ORFC9 antibody (Ab) at a dilution of 1/1000.

Lane 1 is the protein molecular weight marker.

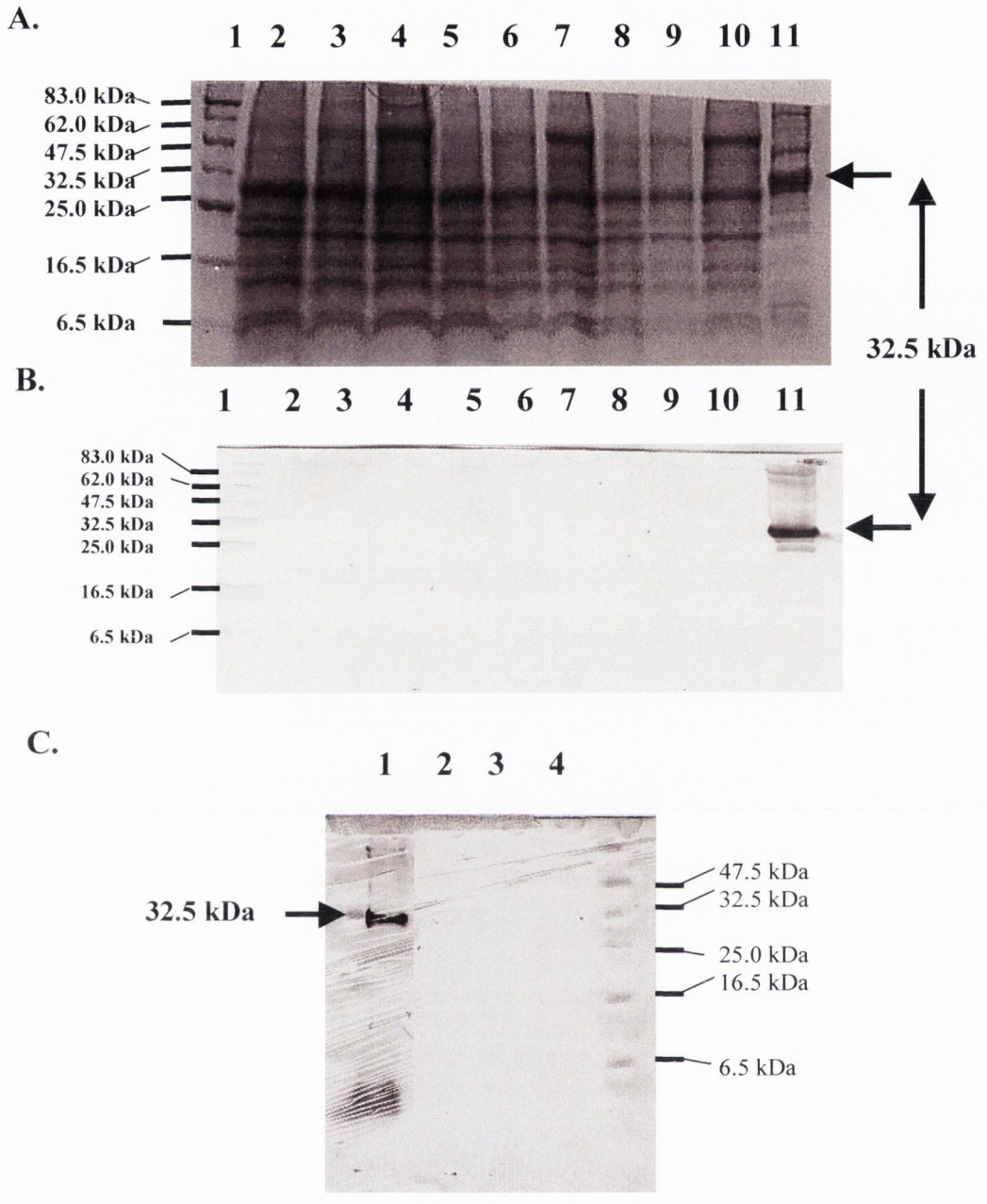


Figure 3.13 Western blot analysis of ORFC9 protein in *L. perenne.*

A. Is a 14 % PAGE coomassie stained gel. Lanes 2 (50 µg), 5 (38µg) and 8 (13 µg) is total leaf protein in the WT line (LPS21), lanes 3 (50 µg), 6 (38 µg) and 9 (13 µg) is total leaf protein in Restored line (Rst1), lanes 4 (60 µg), 7 (50 µg) and 10 (25 µg) is total leaf protein in the Sterile line (CMS9B290). The control lane 11 is the insoluble protein from *E.coli* (**IP**). **B.** Is a western blot of Figure 3.13 A, being probed with anti-orfC9 antibody. **C.** Is an enriched mitochondrial leaf preparation in *L. perenne* run on a 14% PAGE; lane 1 is the recombinant **IP** protein, lanes 2, 3, 4 are the Fertile Wild Type (FtW1), Restored (Rst1) and Sterile (CMS3) enriched mitochondrial protein samples respectively, each lane contains 80 µg-100 µg of protein. Lane 1 is Protein standard used is the pre-stained NEB marker (6-83 kDa).

CHAPTER 4

Identification of an integrated linear plasmid-like element in the mitochondrial genome of a CMS line of *L. perenne*

4.1 Introduction

4.1.1 Linear mitochondrial plasmids in plants and fungi

Mitochondrial plasmids were generally assumed to be circular, until almost 20 years ago, when linear elements were first reported, and were described in both pro- and eukaryotes. Two types of linear plasmids were identified: hairpin plasmids, having covalently closed ends which are found in the human pathogen *Borrelia* (reviewed in Girons *et al.*, 1994) and in the mitochondria of the plant pathogenic fungus *Rhizoctonia solani* (Hashiba *et al.*, 1984; Miyashita *et al.*, 1990), and more commonly, linear plasmids with a covalently-bound protein at the 5' termini (the terminal protein (TP)), found in bacteria, the yeast cytoplasm, the mitochondria of filamentous fungi and, occasionally, the mitochondria of higher plants.

In bacteria, linear plasmids are commonly found in the actinomycetes e.g. *Rhodococcus* (Kalkus *et al.*, 1993; Dabrock *et al.*, 1994; Kesseler *et al.*, 1996) and *Streptomyces* (Kinashi *et al.*, 1987; Wu and Roy 1993; Crespi *et al.*, 1994) with plasmid sizes ranging from 400 kb to 11.7 kb, and are thought to confer advantageous traits on their hosts. The linear DNA plasmids of yeast e.g. *Kluyveromyces lactis* (Gunge *et al.*, 1981), *Pichia acaciae* (Worsham and Bolen, 1990), *Saccharomyces kluyveri* (Hishinuma and Hirai, 1991) replicate in the cytoplasm with plasmid sizes ranging from 14.2 kb to 7.3 kb. They typically encode a 'Killer toxin' and contain genes that encode a DNA polymerase (Dpo) and the TP. The pRK2 linear plasmid in *Pichia kluyveri* is the only example of a mitochondria-associated linear plasmid in yeast and its structure is similar to that of mitochondrial linear plasmids in filamentous fungi (Blaisonneau *et al.*, 1999).

Linear plasmids were originally detected in the mitochondria of plants and fungi by agarose gel electrophoretic analysis of total mitochondrial DNA, which revealed the presence of double-stranded DNA species with a molecular weight much lower (5 – 12 kb) than the main mitochondrial genome (typically greater than 300 kb). They are defined by terminal inverted repeats (TIRs) and a TP covalently attached at the 5' termini. Transcription and replication is typically initiated within the TIRs by, respectively, a phage-like single subunit RNA polymerase (Rpo) and a plasmid-encoded Dpo of the B-type family (Meinhardt *et al.*, 1990). The Dpo and Rpo orfs may either be on the same DNA strand (a head-to-tail arrangement) or on opposite strands (in a convergent tail-to-tail arrangement). Most linear plasmids contain at least one and up to four additional orfs of unknown function.

In contrast to fungi, where they are very common, extrachromosomal linear plasmids have been identified in the mitochondria of only 4 species of higher plants. These include the high copy number S1 and S2 plasmids (Paillard *et al.*, 1985; Levings and Sederoff, 1983) associated with cytoplasmic male sterile (CMS) lines of *Zea mays* (maize) carrying the S cytoplasm (CMS-S or S-mitotype maize); an 11.6 kb linear plasmid in the mitochondria of *Brassica napus* (oilseed rape) (Palmer *et al.*, 1983; Turpen *et al.*, 1987; Handa *et al.*, 2002); a 10.4 kb plasmid in *Beta vulgaris* subsp. *maritima* (sea beet) (Saumitou-Laprade *et al.*, 1989; Michaelis *et al.*, 1997 unpublished data; GenBank Accession No. Y10854); and a 9.2 kb plasmid like sequence recently identified in the FG21 mitotype of *Daucus carota* (carrot) (Robison and Wolyn, 2005). All of these plasmids encode a phage-like Rpo and a Dpo, except for those of maize which encode only one or the other polymerase.

4.1.2 Replication of linear plasmids

Linear plasmids are thought to replicate by a protein-priming mechanism similar to that in the *Bacillus* phage *phi29* (Salas, 1991), where the TP acts as a primer and in association with the Dpo, initiates DNA synthesis in the TIRs. After the intiation step, the same Dpo catalyses chain elongation by strand displacement with the removal of the intiation proteins and the binding of a single stranded binding (SSB) protein to the displaced parental ssDNA (Salas, 1991). Unlike *phi29*, linear mitochondrial plasmids do not encode a recognizable TP which has

led to speculation that this function might be encoded within the Dpo orf. This has recently been confirmed for the pLMP1 linear mitochondrial plasmid of the fungus *Pleurotus ostreatus* (Kim *et al.*, 2000), where the N-terminal region of the Dpo acts as the TP. The identification of a TP-like domain in the Dpo encoded by plant linear mitochondrial plasmids has not so far been reported.

4.1.3 Integration of linear plasmids into the mitochondrial genome

Although linear plasmids do not encode a dedicated integrase function, they can integrate into the mitochondrial genome by recombining with short homologous sequences in mtDNA (Brown and Zhang, 1995). A recent extensive search of the nucleotide sequence databases for plasmid-like sequences in the mitochondrial genomes of filamentous fungi suggests that most linear plasmids can integrate in this way (Cahan and Kennell, 2005). Evidence for the integration of linear plasmids into plant mitochondrial genomes was originally provided by molecular investigations of the maize CMS-S mitotype where integration of the S plasmids is associated with extensive linearization of mtDNA (Schardl *et al.*, 1985). The mitochondrial genomes of *Secale cereale* (rye) (Dohmen *et al.*, 1994) and the liverwort *Marchantia polymorpha* (Weber *et al.*, 1995) were subsequently shown to contain fragments of a linear plasmid-related Dpo despite the apparent absence of homologous extrachromosomal plasmids in the mitochondria of either species. Additional examples of integrated fragments of linear plasmid Rpo- and or Dpo-like orfs were identified in the completely sequenced mitochondrial genomes of *Arabidopsis thaliana* (Unseld *et al.*, 1997), *B. vulgaris* (sugar beet) (Kubo *et al.*, 2000; Satoh *et al.*, 2004), *Oryza sativa* (rice) (Notsu *et al.*, 2002), *Z. mays* (Clifton *et al.*, 2004), *Nicotiana tabacum* (tobacco) (Sugiyama *et al.*, 2005) and *Triticum aestivum* (wheat) (Ogihara *et al.*, 2005). Although not yet completely sequenced, the mitochondrial genome of *D. carota* mitotype SW3 has been shown to contain an integrated but truncated copy of the 9.2 kb extrachromosomal linear plasmid originally identified in carrot mitotype FG21 (Robison and Wolyn, 2005).

4.1.4 Phenotypic consequences of linear plasmid integration

Linear plasmids, by virtue of their ability to integrate into the mitochondrial genome, are potential insertional mutagens. In the fungus

Neurospora, for example, the integration of either of two linear plasmids, pKALILO (Chan *et al.*, 1991) and pMARANHAR (Court and Bertrand, 1992) results in abnormal physiology and a premature senescence or 'death' phenotype. In plants, episomal and or integrated linear plasmids are frequently associated with CMS phenotypes. However, evidence from maize (Escote *et al.*, 1985), *Brassica* (Kemble *et al.*, 1986) and sugar beet (*B. vulgaris*) (Satoh *et al.*, 2004) suggests that this association is fortuitous rather than causative.

4.1.5 Transmission of linear mitochondrial plasmids

Like the mitochondria in which they reside, linear mitochondrial plasmids are, in general, maternally inherited. Exceptionally, however, the *Brassica* 11.6 kb plasmid although showing maternal inheritance, can also be transmitted to progeny through pollen (Erickson *et al.*, 1989).

Evidence supporting the horizontal transfer of linear plasmids has been reported for several species of fungi e.g. within natural populations of the genus *Neurospora* (Marcinko *et al.*, 1994; Arganoza *et al.*, 1994), between *Neurospoa* species (Debets *et al.*, 1994) and between unrelated species (Kempken, 1995). The Dpo of the *Kalilo* plasmid is more closely related to that of the *Ascobolus immerus* pAI2 plasmid than to that of the *Maranhar* plasmid (Rohe *et al.*, 1992; Kempken *et al.*, 1992). Similarly, phylogenetic disorder for plasmids within the same fungal species has been reported *e.g.* the pFV1 and pFV2 plasmids in *Flammulina velutipes* (Nakai *et al.*, 2000), and the pLMP1 and pLMP2 plasmids in *Pleurotus ostreatus* (Kim *et al.*, 2000), which suggest the individual plasmids (of a pair) originated from different fungi.

Horizontal transfer of linear plasmids between fungi and plants has not been reported and so, remains a possibility. Fungi of the genera *Epichloe*, *Neotyphodium* and *Acremonium* are seen as mutualistic symbionts of grasses (especially of *Festuca* and *Lolium* spp.), which during the asexual phase of their life-cycle grow in close association with developing floral organs including ovules, a strategy that ensures their transmission via seeds (Schardl *et al.*, 2004). *Epichloe typhina* is known to contain linear mitochondrial plasmids (Mogen *et al.*, 1991; Chung *et al.*, 1996). It is not inconceivable that such a close association between fungal mycelia and plant reproductive organs might also facilitate rare horizontal transmission of genetic information between the symbiont and its host.

Horizontal transfer between fungi and higher plants is not without precedent. The best documented case involves a mitochondrial homing group I intron of fungal origin which invaded the *coxI* gene of flowering plant mitochondria on more than 1,000 occasions during angiosperm evolution (Vaughn *et al.*, 1995; Cho *et al.*, 1998).

4.2 Chapter Aims

The LpCMSi 9.6 kb region which contains the chimeric gene *orfC9* is uniquely associated with CMS in *L. perenne*. In the previous chapter, I investigated whether *orfC9* might be causally involved in the CMS phenotype. In this chapter, I sequenced the entire LpCMSi region located downstream (3') of *orfC9* with the aim of identifying additional ORFs that might be implicated in the CMS phenotype. Sequence analysis revealed the presence of 3 ORFs, two of which shared extensive homology with Dpos and Rpos encoded by linear mitochondrial plasmids of fungi and higher plants. In addition, sequences highly homologous to LpCMSi were identified at very low copy number in fertile lines of *L. perenne,* some of which were coincidentally infected with fungal endophytes.

4.3 Materials and Methods

4.3.1. Plant material

The *L. perenne* male sterile line CMS-St (CMS9B290), a restored line Rst1, and the fertile maintainer line LPS21 were described in Kiang *et al.* (1993). The fertile *L. perenne* lines FtIp1, FIp2, FIp3, FIp4, FIp5, FIp6, FIp7, FIp8, FIp9 were originally isolated from Irish pastures (V. Connolly, Teagasc); the fertile *L. perenne* line FNz is a commercial endophyte-infected line from New Zealand.

4.3.2 Molecular techniques

General materials and methods used are described in Chapter 2.

4.3.3 Sequencing of clones from CMS-St and the fertile line LPS21

The clones pCMS45, pCMS133, pCMS60, pCMS29, pCMS74 (Kiang and Kavanagh, 1996) which cover the rearranged region in the mtDNA of CMS-St were used to generate sequencing sub-clones and templates.

The clone pLMF56 contains a mtDNA fragment from *L. perenne* LPS21 that contains the LpCMSi insertion site locus.

4.3.4 Primers used in sequencing and analysis of *L. perenne* mtDNA

4.3.4.1 Oligonucleotide primers used in the sequencing of pCMS subclones.

(a) Primers used to sequence subclone pCMS23:

Primer	Sequence (5'-3')
23f.1	GATTCTACTCTAAAACAAGATGCT
23f.2	GAGGAAAATGATATTCCAATCAGT
23f.3	GTAGTCATGATTCTGATCAATTTG
23f.4	TCTGGCGTATATCAACATAAATGA
23f.5	GAGATTTCATGAGATCGCTATTC
23f.6	GAAAACTTTACTGAAGATACTGATATC
23f.7	GATTGCTTGCTCCATCAATTAAG
23f.8	GGTCATAATAGGGATTATGCAAC
23f.9	CTCTGTCATTTATGACTCTACAAG
23f.10	CTGTTCATGATAACTTTATAACAACC
23r.1	GTGGTATTTTATACGCATGTT
23r.2	GTCTTGCAATAGTTTTGTATTGTTTTAAAC
23r.3	CAGTATTGATTATATTCAATGCGTG
23r.4	TAGGGATATGCTTCATGGTATATG
23r.5	GGTTGAAATCATTCATTTTAGCTTG
23r.6	GAATCTAGCACTACTATAGAGATC

23r.7	CAGAATCATGACTACTTATGAGAG
23r.8	CTGATTGGAATATCATTTTCCTCG
23r.9	CATCTTGTTTTAGAGTAGAATCTTG

(b) Primers used to sequence subclone pCMS4:

Primer	Sequence (5'-3')
4f.1	CAGTTTCTAGTTTGTCGTAACATG
4f.2	GAATGACTTGACTCAAACATATCC
4r.1	CGCTCAACTTATCAACATAGATAG
4r.2	GGATATGTTTGAGTCAAGTCATTC

(c) Primers used to sequence subclone pCMS7:

Primer	Sequence (5'-3')
7f.1	GCGGAGAAAACGTCATTTAAC
7f.2	GGTATAAGATGAAATCTGATGTAATG
7r.1	GGTCATGATAAGATGGATATC
7r.2	CAGAAAATGATCAAGATAGTGCATC

(d) Primers used to sequence subclone pCMS11:

Primer	Sequence (5'-3')
11f.1	GATAAGTTCACTAAAAGCTGTGCT
11f.2	CTTGGTTGTATTCGTTGATGTTTAC
11f.3	CAACACCAGTTTGAATTACTGATAC
11f.3.1	GTGAATTTCTCAATTGTGATTGATC
11f.4	GAATCAAATATGGATTTCACCTTT
11r.1	CTATCGGATCGTAAATCGAATATG
11r.2	CACTGGAAGTAACAACAATATTC
11r.3	CATATTACATATCTCCTTACAAAC
11r.3.1	GATGAGATTAGATCAATCACAATTG
11r.4	GAGCAAACCAATGATTACCATTTC

(e) Primers used to sequence subclone pCMS21:

Primer	Sequence (5'-3')
21f.1	TGATTGTGAGTAATTCTAGTTTAGG
21f.2	GTCAAAGGTCGATTTCTTCTATTC
21f.3	CACTAATTTTACGAGACGAAGTAG
21f.4	GTTCCTGTAATGAAACTCATTAC
21f.4.1	CGAATGACAAATGGATGGACTTC
21f.5	CCATCAATGTATTCAGTGATTCG
21f.6	GACAAGAATTCAGATGAATACATG
21f.7	CTGATATGGAAACAATCCAGTTAG
21f.8	CCAATAGTCTATTTTCATAACTTGTC
21f.9	CGAGAATTAAGCAAAGCAAATATG
21f.10	GACATCTCGACAAAGAAAACAATC
21r.2	GTCATCATAGTATTCCATTCGGAAT

21r.3	CAACTCAGGGCATATTTGCTTTG
21r.4	CTATTGGTGTCATTCTCAGCTTC
21r.5	CACTGATTCATCAATCTGAATTCG
21r.6	CATGTATTCATCTGAATTCTTGTC
21r.7	CAATGTCATGAATAGGTATGATTTTC
21r.8	GAATTGCTATGAAAATGAGGTCG
21r.9	CATGTTTGAATTGATCTAATTCTCC
21r.10	GGTTCAAGTACTTTTAGCTAGTATC

(f) Primers used to sequence subclone pCMS10:

Primer	Sequence (5'-3')
10f.2	GACCTCAAATATGCAAAGAGTATT
10f.3	GTTCTGAACAAAGAACAAACAGAG
10f.4	GGATCAAAATGTATCCATACATATC
10f.5	CATATAGATGAGGTCACGTATAC
10r.2	ATTCGTAGTGAAACATGACTTTTTC
10r.3	CACTTAGGTGATACAAAATATCCT
10r.4	GTATCATACATACTCTCTATATAGG
10r.5	GATATCCACATGTGATAATAACTTC

(g) Primers used to sequence a 2.87 kb region in pCMS74:

Primer	Sequence (5'-3')
29f.1	CTTTACAATTTGTATATCTTCCTCG
29f.2	CTTCTACATCATATTGAAGCTGATG
29f.3	GTTTCTTTCCAATAAGTAAAAGAAGC
29f.4	CGAGCATATTCATCTATTTCTTCAAG
29f.5	GTTTTCACAATTTTCTTCCATTGG
29f.6	CCATGAGTATTCACATCATAAAGTC
29r.1	GTGTCGATAATAAGTAATTTCGATC
29r.2	CGTGATAGATGGATATTAAAAAGAC
29r.4	GATTGAAGTACCACAACCATTTAC
29r.3	CAGATAAGGTTTGCTTTACATGC
29r.5	GGATATTTTATCCAGTATGATAGTC
29r.6	CGTGATAATAAACCAAACCAAAATC
29r.7	GTCAGTTCATGTTAAGTTCAAACAC
29r.8	GAGGAAGATATACAAATTGTAAAGC
29F	GATCGAAATTACTTATTATCGACAC
29F1	CACTCACAACGGATATCATTATG
29R1	CGGGAATAAGTAAATTAGATGATAC

(h) Primers used to sequence a 3.84 kb region in pCMS133:

Primer	Sequence (5'-3')
133F	CTCAAACGCTGAGAGAGATTACTATAG
133F1	GTAATTGTATTCTGATTGCATTTCTC
60R1	CTTTAATAAGCGAAGCAACGAG
133R2	CAAACAAACTTACAGTAAGCC

(i) The following primers were used to generate overlap sequence between the clones pCMS45, pCMS11, pCMS74, pCMS4, pCMS21, pCMS60:

Primer	Sequence (5'-3')
11f.1ro	CTCCAACATACAGCATTTTTAGGATG
45.1ro	CTATCGGATCGTAAATCGAATATG
45.2ro	CTTTCTATTAAACAAGGTGCTGTC
7f2.1o	GATGGGTTTATGATCCAGATAGTAC
74.29f	GAGTATCATCTTGATTTTGAGTAG
10rf2	AATGGATTGATACCAAACCAGTTC
10f2r	ACATTTCACTAATGCTTGAACAAAG
4f.1	CAGTTTCTAGTTTGTCGTAACATG
21r.10	GGTTCAAGTACTTTTAGCTAGTATC
pCMS60R	GTGAGAATGATTCCGAGTCTTC

4.3.4.2 Oligonucleotide primers used in the sequencing of pLMF12 subclone

Primer	Sequence (5'-3')
12f1	GAAAAGACCTAGTACTTGTTTACC
12f2	CAAATACGCATATACTGACCTAC
12f3	CCATTATAAGAAATCAGCAACTCG
12f4	CTGTTACCCTTGTATGGTTAATAG
12f5	CAATCGTAAGTTGATTCAAGTCAG
12r1	CGTAAATCTTTCTAATCAGTACGG
12r2	GCCTTCACCCAATTTAATAAATGAG
12r3	CTTTGTTACTATTGCTCGATCTAG
12r4	GATTCAACGACAACAAGGACTAG
12r5	GTCAAGGTAAACAAGTACTAGGTC
12f1.r	CAATTTACAGAGTATTCTTGCATTC

4.3.4.3 Primers used in the amplification of Dpo-like, Rpo-like and intergenic sequences from fertile lines of *L. perenne*

Primers were based on ORF1 (Dpo), ORF2 (Rpo) and ORF3 sequences of LpCMSi

Primer	**Fragment**		**Sequence (5'→ 3')**
N Ter	**DF1**	f	CTAGAACATCTATAGTGACAGATG
TpR1		r	GTACTTATCCTTCTGTATACTGATTG
TpF	**DF2**	f	CTCACTGAACTCATGATAAATGTTC
ExoII		r	GTACTCGTCATTGTGCTTAGTAAAC
TpF1	**D1**	f	CAATCAGTATACAGAAGGATAAGTAC
ExoII		r	GTACTCGTCATTGTGCTTAGTAAAC
ExoII	**D2**	f	CACCAATAGTCTATTTTCATAAC

PolIII		r	GTCCGTATAATAGCAGTCATC
PolIII	**D3**	f	GATGACTGCTATTATACGGACAC
C ter		r	GGAAAGGTATAAGATGAAATCTGATG
Pol IIa	**D4**	f	GTTATTGCTAAATTCATTATATGGTAG
Pol III ter		r	GTATACGTGACCTCATCTATATG
Rpo II	**R1**	f	CAGATGTTTGATTACTTAGCGAAG
Rpo V		r	GATAATTTCTACCACGGAAATCAAG
Rpo V	**R2**	f	CTTGATTTCCGTGGTAGAAATTATC
Rpo X		r	CTGATGGATGAAGTTCGCAAAG
Rpo FV	**R3**	f	CGTTCAGAAAGCCATGTTTGAACAAG
Rpo RV		r	CAATCCAACTGATCTTGGGTTAACTC
Rpo XF	**R4**	f	CAATCGGTCTGCATGATCACTAC
Rpo X R		r	GAAACTGAGCGTAATCAGCATCAC
Rpo X	**R5**	f	CTTTGCGAACTTCATCCATCA
Rpo C ter		r	CACTTTTGAATGAACACTATACTCG
N Ter	**R6**	f	GAGTTACTAGAGTCGACTCTAATGTTG
N R1		r	CCGCAATTCATACTGAAATCATCAATTTG
Rpoa (NF1)	**R7**	f	CTTCTGGGATACACTTTTAGCTAGTG
Rpolb (II)R		r	CAAACATCTGTGTCCCTATAGCATAAG
Jl F	**JL**	f	CGAAATATCTCTCTGTCTTGTTCCG
Jl R		r	GCACCTTTTCTTTGGAATTGTCATTCC
Jr F	**JR**	f	CAACATTAGAGTCGACTCTAGTAACTA
Jr R		r	GAGTTACCAATCCGTGTCGATAATAAG
orfC9F	**O**	f	GAAACAATTAAGTATTTACTCG
orfC9R		r	GCAACAACAACTTTACTTATTC
DNATpF1	**ORF1/2a**	f	CGTGCAATCAGTATACAGAAGGATAAGTACC
RNA TerR		r	CTGAGTTACTAGAGTCGACTCTAATGTTGA
ORF1F	**ORF1/2b**	f	CTCAGAAGTCCTGATACTGTCAC
ORF3R2		r	CTAGTTTGTCGTAACATGTTCAACG

The following primers were used in the bidirectional sequencing of the PCR fragments generated by the primer pairs listed above. Overlap fragments used for sequencing purposes are highlighted below.

Fragment (size bp)	Primer	Sequence (5'-3')
D1 (738)		
	TpF1	CAATCAGTATACAGAAGGATAAGTAC
	ExoIF	CTGATATGGAAACAATCCAGTTAG
	ExoIIR	GTACTCGTCATTGTGCTTAGTAAAC
	ExoIIR1	CTCTAACTGGATTGTTTCCATATCAG
D2 (1,295)		
	ExoIIF	CACCAATAGTCTATTTTCATAAC
	ExoIIF1	CTCGAGAATTTAGCAAAGCAAATATGC
	PolIF	GTCTATGATGTGAATTCATTATATCCATC
	PolF1	CACTGAGGACCTCAAATATGC
	PolIIaF	GTTATTGCTAAATTCATTATATGGTAG
	PolIIIR	GTCCGTATAATAGCAGTCATC
	PolIIaR1	CTACCATATAATGAATTTAGCAATAAC
	PolR1	GCATATTTGAGGTCCTCAGTG
	Pol R1	GATGGATATAATGAATTCACATCATAGAC
	ExoIIR	GCATATTTGCTTTGCTAAATTCTCGAG
D3 (703)		
	PolIIIF	GATGACTGCTATTATACGGACAC
	PolIIIF1	CATATAGATGAGGTCACGTATAC
	CterR	GGAAAGGTATAAGATGAAATCTGATG
	CterR1	GTATACGTGACCTCATCTATATG
D4 (555)		
	Pol IIa	GTTATTGCTAAATTCATTATATGGTAG
	Pol III ter	GTATACGTGACCTCATCTATATG
R1 (843)		
	Rpo IIF	CAGATGTTTGATTACTTAGCGAAG
	Rpo IIIF	CACTGTCAAGATTCGAAGTATATC
	Rpo IVF	GTCAACTTCCTGAAGTATTTACAG
	Rpo IIIR	GATATACTTCGAATCTTGACAGTG
	Rpo IVR	CTGTAAATACTTCAGGAAGTTGAC
	Rpo VR	GATAATTTCTACCACGGAAATCAAG
R2 (1,022)		
	Rpo V F	CTTGATTTCCGTGGTAGAAATTATC
	Rpo VF1	GAGTTAACCCAAGATCAGTTGGATTG
	Rpo VIIF	CAAGATGCATCAGCTAGTGCATATCAGATC
	Rpo IXF	CAATCGGTCTGCATGATCACTACG
	Rpo VIIR1	GATCTGATATGCACTAGCTGATGCATCTTG
	Rpo VIIR	GATATGCACTAGCTGATGCATCTTG
	Rpo XR	CTGATGGATGAAGTTCGCAAAG
R3-overlap (241)		
	Rpo FV	CGTTCAGAAAGCCATGTTTGAACAAG
	Rpo RV	CAATCCAACTGATCTTGGGTTAACTC

R4 (502)

Rpo XF	CAATCGGTCTGCATGATCACTAC
Rpo X R	GAAACTGAGCGTAATCAGCATCAC

R5 (808)

Rpo X	CTTTGCGAACTTCATCCATCA
Rpo C ter	CACTTTTGAATGAACACTATACTCG

R6 (800)

NTerF	GAGTTACTAGAGTCGACTCTAATGTTG
NPolF	GCCTTCTGGGATACACTTTTAGCTAG
NPolR	CACTAGCTAAAAGTGTATCCCAGAAGGC
NR1	CCGCAATTCATACTGAAATCATCAATTTG

R7 (836)

Rpoa (NF)F	CTTCTGGGATACACTTTTAGCTAGTG
Rpo (NF1)F	GATTTCCATCAGCTTCAATATGATG
Rpo IF	GATGTCGGATCCAGATTATATTAAG
Rpo (NF1)R	CATCATATTGAAGCTGATGGAAATC
Rpo IR	CTTAATATAATCTGGATCCGACATC
Rpob (II)R	CAAACATCTGTGTCCCTATAGCATAAG

R8-overlap (594)

Rpo IF	GATGTCGGATCCAGATTATATTAAG
Rpo IIIR	GATATACTTCGAATCTTGACAGTG

4.3.4.4 Primers used to amplify fragments from cDNA templates of sterile (CMS9B290) and restored (Rst1) lines.

Primer	Fragment		Sequence (5'→ 3')
ExoII	**DP1**	f	CACCAATAGTCTATTTTCATAAC
Pol I		r	GATGGATATAATGAATTCACATCATAGAC
Pol I	**DP2**	f	GTCTATGATGTGAATTCATTATATCCATC
Pol III		r	GTCCGTATAATAGCAGTCATC
ORF1	**DPN1**	f	CTCAGAAGTCCTGATACTGTCAC
orf3Ra		r	CAACACAATAGGCTAAGGTTATCCAC
orf3Fa	**N1**	f	GATCAAGATAGTGCATCCAATGATGG
orf3Rb		r	GATATCCTTCAACCACTGGTGATTC
Rpo II	**RP1**	f	CAGATGTTTGATTACTTAGCGAAG
Rpo V		r	GATAATTTCTACCACGGAAATCAAG
Rpo V	**RP2**	f	CTTGATTTCCGTGGTAGAAATTATC
Rpl X		r	CTGATGGATGAAGTTCGCAAAG

4.3.4.5 Primers used for generating protein expression constructs

Primers used to amplify a 2.89 kb fragment of Dpo for cloning into the pTwin-IMPACT expression vector (*NEB*).

pTwNcoF 5' GGTGGT**CCATGG**TGAACACTAATTTTACGAG 3'

pTwBamH1R 5' GGTGGT**GGATCC**TCAGTCTGAGTCTGACTCATC 3'

Primers used to confirm sequence at the N-terminal junction and C-terminal junction of 2.9 kb insert in the pTwinLpDpo construct

IF 5' ACTGGGACTCCATCGTTTC 3'

10rf2 5'AATGGATTGATACCAAACCAGTTC3'

Primers used to amplify a 843bp fragment of Dpo cloned into pET32a expression vector (*Novagen*).

TpF 5' CTC**GGATCC**ATGGAGGATAAGAAGGACAAAAAATAC3'

EIIIR 5'GCGT**AAGCTT**CTCTTGAGGAAGACAATAAATATCACC 3'

Primers used to confirm sequence of the 843 bp fragment in pET32LpDpo construct

StagF 5' GCTGCTGCTAAATTCGAACGCCAG 3'

ExoIIF 5' CACCAATAGTCTATTTTCATAAC 3'

Pol IR 5' GATGGATATAATGAATTCACATCATAGAC

T7R 5' GCTAGTTATTGCTCAGCGGTGGC 3'

4.3.5 Nucleotide sequencing and sequence analysis

DNA sequencing reactions were carried out according to manufacturers specification as outlined in Chapter 2. Sequencing was performed by primer walking across mitochondrial clones (Kiang and Kavanagh, 1996) and the generated subclones. Sequence homologies were identified using the BLAST search algorithm (www.ncbi.nlm.gov/blast/; Altschul *et al.*, 1997). Sequence alignments and manipulations were performed using Gene Jockey II (Biosoft) and ClustalW (www.ebi.ac..uk/clustalw/; Thompson *et al.*, 1994). Phylogenetic trees were constructed using the Mega-Align program 'Molecular evolutionary Genetics Analysis version 2.1' (www.megasoftware.net/; Kumar *et al.*, 2001) and editing of sequencing data was performed using the Gene Doc program (www.psc.edu/biomed/genedoc).

4.3.6 RNA analysis

Total RNA (9 μg) extracted from a wild-type fertile line (LPS21), the sterile line CMS9B290 and the restored line (Rst) of *L. perenne* was used in northern blot analyses. Blots were probed separately with 2.9 kb (orf1) and 4.5 kb (orf1, orf2, orf3) fragments. Reverse transcription (RT) PCR was used to generate cDNA from 1-5 μg of DNAase-treated total RNA from sterile and restored lines by using the reverse transcriptase Superscript II (General Materials and Methods Chapter 2).

4.3.7 Protein expression

Recombinant protein expression methods and the conditions used are described in Chapter 2. The pET32a vector (www.novagen.com) and the pTwin-IMPACT vector system (www.neb.com) were used to express different regions of Dpo according to the manufacturers instructions.

4.3.8 PCR conditions

4.3.8.1 PCR amplication using total cellular DNA template

Total DNA was isolated from 1 g of fungal infected lines of *L. perenne,* and PCR amplification was performed on these samples as follows: 10 ng of template DNA was amplified in a 50 μl volume using 0.5 units *Taq* DNA polymerase. The denaturation step was at 94°C for 1 min., annealing step 60°C for 1 min., polymerisation step at 72°C for 1 min. Twenty five cycles were performed with a final extension at 72°C for 5 minutes.

4.3.8.2 PCR amplication of fragment from orf1 for protein expression

For protein expression work, PCR amplification of regions from *orf1* was performed on 10 ng of pCMS60, and was carried out in a 50 μl volume using 1 unit of *Pfu* DNA polymerase. The conditions used were: Denaturation at 94°C for 1 min., annealing at 62°C for 1 min., polymerisation step at 72°C for 1 min. Thirty cycles were performed with a final extension at 72°C for 5 minutes.

4.3.8.3 PCR amplification from the cDNA template

PCR amplification from first strand cDNA template prepared from sterile and restored lines was performed in a 50 μl volume using 0.5 units of *Taq* DNA

polymerase. The conditions used were: Hot start at 95°C for 1 min., denaturation step at 94°C for 1 min., annealing at 60°C for 1 min., polymerisation step at 72°C for 1 min. Thirty five cycles were performed followed by a final extension at 72°C for 10 min.

4.3.8.4 Long range PCR amplification from total DNA template

Long range PCR amplification was attempted on 50-100 ng of total cellular DNA extracted from *L. perenne* using the 'Expand Long Template PCR system' (LaRoche). Reactions were performed by initially designing two separate master mixes. Master mix 1 contained dNTPs (10 mM), primers (50 ng/µl), template DNA (50-100 ng/µl), 1X Expand long buffer and sterile water, making up a final volume of 25µl. Master mix 2 contained 1X Expand buffer, 3.75 Units long range enzyme made up to a volume of 25 µl using sterile water. The long range PCR reaction was carried out on the combined master mixes under the following conditions using a PTC-100 machine: - [95.0°C for 2 min., 95.0°C for 30 sec., 60.0°C for 30 sec., 72.0°C for 40 sec.,] - for a 10 cycle period, - [95.0°C for 30 sec., 60.0°C for 30 sec., 72.0°C 40 + 20 S/cycles] for a 30 cycle period, and a 72.0°C extension for 7 minutes.

4.3.8.5 Proteinase K treatment of total cellular DNA.

Leaf tissue of *L. perenne* was initially ground to a powder in liquid nitrogen. Total DNA was extracted from powdered tissue according to the manufacturers specifications using the Puregene Plant DNA kit (Gentra Systems ©). Extraction was performed in the absence of phenol. Total DNA was treated with proteinase K (400 µg/ml)) and incubated at room temperature for 18 hr.

4.4 Results

An RFLP-based analysis of the mitochondrial genome of *L. perenne* CMS-St identified a novel 9.6-kb region, designated LpCMSi, located 89 bp downstream of the *atp9* gene, which was absent in fertile lines and was therefore diagnostic of the CMS cytoplasm (Kiang *et al.*, 1993; Kiang and Kavanagh, 1996). This region and approximately 3.6-kb of flanking mtDNA is contained in three contiguous restriction fragments (from left-to-right, relative to *atp9*): 6.0-kb *Hind*III, 4.5-kb *Hind*III/*Bam*HI, 2.9-kb *Bam*HI/*Hind*III (Figure 4.7), which were cloned in the vector pUC19 to give, respectively, pCMS60, pCMS45 and pCMS29 (Kiang and Kavanagh, 1996). These plasmids were used in the present study as a source of hybridization probes and in the preparation of subcloned fragments for use as LpCMSi sequencing templates.

4.4.1 Construction of subclones spanning LpCMSi and the homologous insertion site locus in the fertile *L. perenne* line LPS21

A restriction map of pCMS45 was prepared using restriction enzymes corresponding to the multiple cloning site of pUC19 in order to generate a series of sub-clones covering the region for use as sequencing templates (Figure 4.1). The restriction fragments of interest are the *Bam*HI/*Xba*I 2.3 kb (Figure 4.1A, lane 2 and Figure 4.1C, lane 2), *Xba*I/*Sal*I 1.1 kb (Figure 4.1A lane 6, and Figure 4.1B lane 2), *Hind*III/*Sal*I 700 bp (Figure 4.1A lane 5, and Figure 4.1C lane 3) and *Bam*HI/*Xba*I 400 bp (Figure 4.1A lane 2, and Figure 4.1C lane 2) which span the entire 4.5 kb fragment. These sub-fragments were gel purified and cloned into pUC19 to give the subclones pCMS23, pCMS11, pCMS7 and pCMS4, respectively. Subclones containing specific fragments were identified by restriction digest analysis (Figure 4.2).

The 1.67 kb *Xba*I to *Sma*I[1] region in pCMS60 had been sequenced previously (Kiang and Kavanagh, 1996). To completely sequence this clone, subclones for the *Sma*I[1] to *Hind*III[2] region of the CMS-associated portion of pCMS60 were generated (Figure 4.3). Digestion with *Hind*III and *Sam*I gave two sub-fragments i.e. a *Sma*I[1]/*Sma*I[2] fragment (2.1 kb) and *Sma*I[2]/*Hind*III[2] fragment (1.136 kb). The fragments were gel purified and cloned into pUC19 to give

pCMS21 and pCMS10 respectively which were verified by restriction digestion (Figure 4.4).

A third clone pCMS74 (Kiang and Kavanagh, 1996) was digested with *Bam*HI/*Hind*III to give two fragments of 4.5 kb and 2.9 kb (Figure 4.5A). The 4.5 kb product corresponds to the CMS-diagnostic fragment previously cloned as pCMS45 (Kiang and Kavanagh, 1996). The 2.9 kb *Bam*HI/*Hind*III fragment contains the right junction site JR, followed by the homologous mtDNA region shared between the fertile (LPS21) and sterile *L. perenne* line (CMS9B290) as illustrated in Figure 4.7. Restriction digest confirmation of pCMS29 is shown in Figure 4.5B.

The clone pLMF56 spans the region in fertile (LPS21) mtDNA that contains the insertion site locus. Earlier studies had sequenced a portion of this region (Kiang and Kavanagh, 1996). Digestion of pLMF56 with the restriction enzymes *Eco*R1 and *Sma*1 (Figure 4.6) yielded fragment sizes of 4.0 kb, 3.0 kb and 1.2 kb (lane 5). By comparing the digest pattern with a detailed restriction map of the region (Figure 4.7), it was concluded that the 1.2 kb fragment contains mtDNA sequences located downstream of the insertion site that are absent in sterile mtDNA (CMS2B90). This region of the insertion site locus was named the insertion locus non-homology region. The corresponding 1.2 kb fragment was cloned into pUC19; confirmation of this clone is shown in Figure 4.6B.

Figure 4.7 summarises the structure of the LpCMSi region in CMS9B290 relative to the insertion site locus in LPS21. LpCMSi is defined by left and right junctions (JL and JR, respectively) with mtDNA sequences. JL comprises an 'AT' rich motif containing a partial repeat (5' ATAATATATAATATAATA 3'). JR comprises an inverted repeat present in both CMS9B290 and LPS21 (5' ACTTCAATCGTAAGTTGATTCAAGT 3') that is predicted to form a stem loop structure (Kiang and Kavanagh, 1996). Sequences outside JL and JR are present in the mtDNA of both sterile and fertile lines. As identified by previous workers, the insertion locus non-homology region, present in fertile but deleted in CMS mtDNA, contains an open reading *orfL* located downstream and immediately adjacent to *atp9*. In CMSi, a chimeric orf, orfC9 is located in the same relative position downstream of *atp9*. The *atp9* gene therefore defines a region of homology shared by CMS9B290 and LPS21 immediately upstream of JL (Figure 4.7).

The oligonucleotide primer walking strategy was used to sequence the various subclones spanning the CMSi region in sterile mtDNA and the corresponding insertion site locus in fertile mtDNA is illustrated in Figures 4.8, 4.9 and 4.11. The nucleotide sequences of each region is presented in Figure 4.10 and 4.12.

4.4.2 Identification of open reading frames in LpCMSi

LpCMSi possesses an overall GC content of 36%, lying within the 30 – 39% range reported for episomal mitochondrial linear plasmids in plants, but significantly lower than that reported for the main mitochondrial genome (43-45%) (Handa *et al.*, 2002; Clifton *et al.*, 2004).

Sequence analysis revealed the presence of 4 orfs within LpCMSi. The smallest is the chimeric gene *orfC9*, referred to above. This ORF comprises the first six codons of *atp9* followed by 117 codons of unknown sequence. Downstream of orfC9, there are three ORFs each with methionine (ATG) as the start codon (Figure 4.13). ORF1 (2.97 kb) has a coding capacity of 1,002 amino acids (molecular mass of 111 kDa), ORF3 (1.095 kb) encodes 365 amino acids (molecular mass of 40.51 kDa), and on the opposite strand ORF2 (3.855 kb) encodes 1,285 amino acids (molecular mass of 140.6 kDa) (Figure 4.13). The single TGA codon found in ORF1 and ORF3 at positions 966 and 349, respectively, is expected to be translated as Trp (rather than a stop codon) as has been utilised for fungal and plant mitochondria (Griffiths, 1995; Kim *et al.*, 2000; Robison and Wolyn, 2005). Figure 4.14 compares the arrangement of the ORFs in LpCMSi with that in various episomal mitochondrial linear plasmids of plants and fungi.

4.4.2.1 ORF1 encodes a family B-type DNA polymerase

The predicted ORF1 translation product was submitted to the GenBank, EMBL and DDBJ databases with a view to identifying similar sequences.

ORF1 was significantly similar to members of Dpo family B (protein-primed DNA polymerases) encoded, for example, by linear plasmids, bacteriophages and viruses that infect eukaryotes (Jung *et al.*, 1987). ORF1, hereafter referred to as LpDpo, contains the two enzymatic domains characteristic of family B Dpos (Figure 4.13, Figure 4.15). The first, located at the N-terminal

end (aa residue 246-486), is a proof-reading domain with a predicted 3'-5' exonuclease activity, characterised by the motifs Exo I, Exo II and Exo III (Bernad *et al.*, 1989). The second domain, located at the C-terminal end (aa residue 532-920), is predicted to possess polymerase activity and to bind deoxynucleotide triphosphates and pyrophosphates by means of the highly conserved motifs Pol I, Pol IIa, Pol IIb, Pol III (Bernad *et al.*, 1987; Bois *et al.*, 1999, Handa *et al.*, 2002). Additionally as shown in Figure 4.15, two further polymerase motifs Pol IV, Pol V were identified, but were not as highly conserved (Blanco *et al.*, 1991; Bois *et al.*, 1999).

LpDpo shares, for example, 32% amino acid identity with the putative Dpo encoded by the *Z. mays* S1 plasmid (Paillard *et al.*, 1985), 31% identity with ORF1 of a linear plasmid from *B. vulgaris* (Acc. No. Y10854; unpublished), 29% identity with ORF5 of the *B. napus* plasmid described by Handa *et al.*, (2002), and 27% identity with ORF2 of the linear plasmid recently described from *D. carota* mitotype FG21 (Robison and Wolyn, 2005). LpDpo shared lower levels of similarity with the putative Dpos encoded by fungal linear mitochondrial plasmids e.g. 20% identity with ORF1 of the *Claviceps purpurea* plasmid pCLK1 (Oeser and Tudzynski, 1989).

Conserved motifs characteristic of the N-terminal 3' – 5'exonuclease (Exo) domain and the C-terminal polymerization (Pol) domain, conserved in all proof-reading DNA polymerases were identified in LpDpo, and in highly homologous sequences amplified by PCR from FIp4, a fertile wild *L. perenne* accession isolated from Irish grasslands and FNz, a commercial *L. perenne* cultivar from New Zealand infected with the fungal endophyte *Epichloe typhina* x *Acremonium lollii* (see below, section 4.4.3).

LpDpo in common with the Dpos encoded by other linear plasmids contains a large N-terminal extension in Figure 4.13 (amino acid residues 1 – 245), relative to the family B viral Dpos, that shares low homology with the TP (Terminal protein) of bacteriophage *phi29* (Meijer *et al.*, 2001). The corresponding N-terminal domain in the putative Dpo encoded by pMLP1/2, a mitochondrial linear plasmid of the white rot fungus *Pleurotus ostreatus*, has been shown to function as a TP (Kim *et al.*, 2000). However, the levels of sequence identity shared by, for example, the TP of *phi29* (Hermoso *et al.*, 1985; Jung *et*

al., 1987) and the N-terminal region of LpDpo or pLMP1 were not significant (ca. 12%).

Replication of *phi29* has recently been shown to require a specific interaction between the TP and a conserved region located between the Pol I and Pol IIa domains known as TPR-1 (terminal protein region 1) (Dufour *et al.*, 2000). Site-directed mutagenesis of a conserved aspartic acid/glutamic acid (D/E) residue in *phi29* TPR-1 (D332Y) resulted in defective protein-primed DNA replication. LpDpo contains a homologous TPR-1 motif, corresponding to amino acid residues 600 – 647, with the conserved D/E residue located at position 633 (Figure 4.15).

4.4.2.2 ORF2 encodes a single subunit, phage-like RNA polymerase

A BLAST search of the predicted ORF2 translation product revealed extensive similarity with the single subunit Rpos encoded by bacteriophages and mitochondrial plasmids. The level of sequence similarity was highest between ORF2 (referred to hereafter as LpRpo) and Rpos encoded by plant mitochondrial linear plasmids, ranging from 24.8% identity with ORF1 of the *Z. mays* S2 plasmid (Levings and Sederoff, 1983) to 22.0% identity with ORF6 of the *B. napus* 11.6 kb linear plasmid (Handa *et al.*, 2002). The amino acid sequence of LpRpo and highly homologous sequences from the fertile *L. perenne* lines FIp4 and FNz (see below section 4.4.3) were compared with Rpo sequences from linear plasmid-like elements of fungi and other plant species (Figure 4.16). The single subunit Rpo family is characterized by eleven conserved motifs. Motifs I and II are poorly conserved between different subfamily groupings. However, motifs III – XI are highly conserved between diverse species (McAllistair and Raskin, 1993) and of these, nine are easily recognisable in LpRpo (Figure 4.16).

4.4.2.3 ORF3 is unique

The predicted 40.5 kDa polypeptide encoded by *orf3* was novel, showing no significant similarity to any proteins in the public sequence databases. Such proteins are commonly encoded by linear plasmids and may play a role in their maintenance and or transmission, as suggested by Handa *et al.*, (2002).

4.4.3 PCR detection of CMSi-like sequences in fertile lines of *L. perenne*

Total leaf DNA was isolated from fungal infected lines of *L. perenne* which included a commercial endophyte-infected New Zealand grass line (FNz) (90 – 95% seed infection with *Acremonium lollii*) and 9 fertile *L. perenne* accessions from Irish pastures (FIpn) that showed approximately 10% endophyte infection (supplied by Connolly V, *Teagasc*). Total DNA from these lines was screened by PCR using primer pairs based on LpDpo (primers D1 to D4; Figure 4.17) and LpRpo sequences (primers R1 to R8; Figure 4.19). A set of overlapping fragments corresponding to an entire Rpo orf and all but the N-terminal 134 codons (nucleotides 405-3005) of a Dpo orf, was amplified from total DNA of FIp4 and FNz, and their nucleotide sequences determined by direct sequencing of PCR products. No PCR products were detected in lines FNz, FIp4 and Ft (Wt) when the primer pairs DF1, DF2 were used (Figure 4.17).

Alignment of the LpDpo amino acid sequence with the Dpo sequences from FNz and FIp4 (FNzDpo and FIp4Dpo, respectively) revealed that despite several codon alterations, the reading frame was preserved and the Exo and Pol motifs were highly conserved (Figure 4.18). Furthermore, compared with LpDpo, the nucleotide sequence of FNzDpo and FIp4Dpo differed at 30 and 70 positions, respectively, resulting in 20 (2.3%) and 43 (5%) amino acid substitutions, respectively, in the corresponding orfs (Table 4.1). Thus, FNzDpo is more closely related to LpDpo than is FIp4Dpo. This relationship was also evident in sequence comparisons between the various Rpo orfs: Sequencing of the PCR products in FNzRpo and FIp4Rpo (Figure 4.19) contained 70 and 144 nucleotide substitutions, respectively, resulting in 50 (4%) and 105 (8%) amino acid substitutions (Table 4.2).

Similarly, alignment of the LpRpo amino acid sequence with the Rpo sequences from FNz and FIp4 (FNzRpo and FIp4Rpo, respectively) showed high conservation of the Pol motifs (Figure 4.20). Surprisingly, both FNzRpo and FIp4Rpo contained an in-frame deletion of 12 codons in the region located between the RI and RII domains (Figure 4.20). The region contains 8 almost identical copies of a repeated sequence motif 5'-TACTCTAAAACAAGAT-3', of which 3 copies flank each end of the deleted region, while 2 are included within the deleted region (Figure 4.21). Hence, the deletion event is likely to have been mediated by recombination. Remarkably, none of the nucleotide changes

identified in the FNz and FIp4 Dpo-like and Rpo-like sequences disrupt the reading frames by creating premature stop codons or frameshifts.

Finally, in order to assess whether the Rpo-like and Dpo-like sequences identified in the FNz and FIp4 lines might, like LpCMSi, be integrated at the same location in the *L. perenne* mitochondrial genome, PCR amplifications were performed using the primer pairs JR and JL which span the right and left junctions, respectively, between LpCMSi and mtDNA (Materials section 4.3.3.2, Figure 4.22A) and primers specific for *orfC9*. In Figure 4.22 PCR products of the expected sizes (156 bp, 268 bp and 411 bp, respectively) were amplified only in DNA samples from the CMS-St line suggesting that the Rpo-like and Dpo-like sequences identified in the FNz and FIp4 lines might be of episomal (and hence, extrachromosomal) origin.

4.4.4 Evidence that the LpCMSi-like sequences in fertile lines of *L. perenne* may exist as extrachromosomal replicons

To determine if the LpCMSi-like sequences described above reside extrachromosomally or are integrated in the plant mt genome, Southern blot analyses was carried out on total DNA samples from fertile and CMS lines of *L. perenne* (Figure 4.23), treated (i) with and (ii) without the restriction enzyme *Hind*III, using a 555 bp LpDpo fragment and an 838 bp LpRpo fragment as probes.

In the case of the *Hind*III-digested samples, single bands of the expected sizes, 6.0 kb and 7.4 kb, were detected in the CMS lines following hybridization with the LpDpo probe and the LpRpo probe, respectively. However, no signals were detected in any of the fertile lines examined, indicating lack of integration of CMSi-like sequences in fertile lines (Figure 4.23A, B).

The samples that had not been subjected to *Hind*III digestion were treated instead with proteinase K (400 μg/ml) for 18 hr. This is known to remove TP from linear plasmids, allowing them to migrate during agarose gel (stained with ethidium bromide) electrophoresis according to their true molecular mass. Figure 4.23C shows undigested total DNA which migrates near the top of the gel due to its high molecular weight, but no plasmid-like lower molecular weight bands could be identified in any of the proteinase K-treated samples. Furthermore,

hybridisation with either the LpDpo or LpRpo probe failed to detect any hybridizing bands in either the CMS or fertile samples. Taken together these results suggest (i) neither LpCMSi in CMS lines, nor the LpCMSi-like sequences detected by PCR in fertile lines reside extrachromosomally or (ii) if they do, they exist at a copy number that is below the level of detection by Southern hybridization (Figure 4.23 D, E).

Using another approach to detect potentially integrated LpCMSi-like sequences in fertile lines, PCR was performed using the primer pair Ja and Jb which span the LpCMSi junctions, and a primer pair specific for *orfC9*. Although PCR product of 156 bp, 268 bp and 411 bp respectively were amplified in the CMS-St line, no products were amplified in the FNz, FIp4 or Ft lines (Figure 4.22B). This suggests that LpCMSi-like sequences are unlikely to be integrated into mtDNA of fertile lines, at least not at the same locus as in the CMS-St line.

An attempt was also made to amplify an almost full-length CMSi-like sequence from FNz and FIp4 by using long range PCR. The primer pair ORF1/2a, which anneals to the 5' end of LpDpo to the 3' end of LpRpo, amplified an 8.2 kb fragment in the sterile line but a fragment of a similar size was not amplified in the FNz and FIp4 lines (Figure 4.24A). It was decided therefore, to design a primer pair (ORF1/2b) to amplify only the region containing *orf3* on the basis its sequence would provide the necessary overlap between the Dpo and Rpo sequences obtained for FNz and FIp4. A PCR product of 2.2 kb was amplified in FNz and FIp4, and in the CMS-St and the restored fertile line Rst1 (Figure 4.24B). The PCR products were directly sequenced and the sequence data added to the previously obtained FNz and FIp4 LpDpo and LpRpo sequences resulting in a contig of 8.03 kb (data not shown). The FNz and FIp4 8.03 kb contigs shows 98.64% and 97.4% identity, respectively, to LpCMSi.

4.4.4.1 Verification of endophyte infection in the *L. perenne* line FNz

The *L. perenne* line FNz, a commercial seed cultivar from New Zealand was confirmed to be infected with the hybrid fungal endophyte *Epichloe typhina* x *Acremonium lollii* by PCR amplification (Figure 4.25) and sequencing of the *pyr4-2* (Figure 4.26) and *tub2-2* (Figure 4.27) genes encoding, respectively, orotidine-5-monophosphate decarboxylase and tubulin, using the primer pairs Pyr4F 5' CAAACCCTCAAGGCGACTTACGC 3', Pyr4R 5' CCCGCCGAGG

ACATCTGC 3' (Collett *et al.*, 1995), and Tub2F 5' GGTTCATCTTCAAACC GGTCAGTGCG 3', Tub2R 5' CGACGAGAACAGCGCGAGGAAC 3' (Schardl *et al.*, 1994).

4.4.5 Sequences homologous to ORF1 and ORF2 in plant mitochondrial genomes.

Analysis of the 7 completely sequenced plant mitochondrial genomes indicated that only the species *Z. mays* (NB) (Clifton *et al.*, 2004); *B. vulgaris* (Kubo *et al.*, 2000), *O. sativa* (Notsu *et al.*, 2002) and *T. aestivum* (Ogihara *et al.*, 2005) contain both Dpo-like and Rpo-like sequences (Figure 4.28 and Figure 4.29). Dpo-like sequences were also identified in the *M. polymorpha* (Oda *et al.*, 1992) and *N. tabacum* (Sugiyama *et al.*, 2005) mitochondrial genomes. The *B. napus* genome, in contrast, contained neither Dpo nor Rpo sequences (Handa, 2003).

A Blast search using LpDpo and LpRpo as query sequences identified Dpo- and or Rpo-like sequences (some full-length, others mere fragments) putatively derived from linear plasmid, integrated into the mitochondrial genome of *D. carota* (Robison and Wolyn., 2005), *Z. mays* (Garcia *et al.*, 1988; Acc. no. AY506529), *S. cereale* (Dohmen and Tudzynski, 1994), and *B. vulgaris* ((Acc. no. Z34298) and (Kubo *et al.*, 2000)). In *B. vulgaris* 3.6% of the mt genome is homologous to plasmid sequences (Acc. no. BA000009; Z34298). These sequences mostly consist of disrupted Dpo- and Rpo-like orfs or orf fragments (Acc no. Y10854). Multiple alignments of the various sequences are shown in Figures 4.28 and Figure 4.29. The alignments include mt linear plasmids from higher plants, fungi (*C. purpurea* and *Blumeria gramminis*), and the phage *phi29.*

Sequence comparisons of the various Dpo-like sequences included the exonuclease domains Exo I, II, III, the polymerase domains Pol I, Pol IIa, IIb and the putative Terminal Protein Region 1 domain (TPR1) (Figure 4.28). The Pol III and Pol IV domains were joined together including the inter-domain sequence to form a single large Pol III domain (Figure 4.28) as identified by previous workers (Robison and Wolyn, 2005). However, because of poor sequence conservation the Pol V domain was excluded from this analysis.

With reference to Rpo-like sequence the most conserved domains i.e. III, IV, V, VII, VIII, X and XI, were included in our analyses (Figure 4.29).

4.4.6 Phylogenetic analysis of Dpo- and Rpo-like sequences in plants and fungi

Both episomal and integrated Dpo- and Rpo-like sequences were recovered from higher plants, fungi, bacteriophages and viruses was extrapolated from Genbank/EMBL database. Multiple alignment of these sequences was carried out using ClustalW (www.ebi.ac.uk/clustalw) using only the most conserved regions and sequence in between these conderved domains of the respective proteins, being similar to the approach used by Kempken *et al.*, 1992 (data not shown). This corresponded to the DNA binding domain in Dpo-like sequences (Blanco *et al.*, 1991) which begins 50 amino acids upstream of the conserved first aspartic acid residue (D) in Pol I and ends approximately 50 amino acids downsteam of the conserved DTDS domain in Pol III (Kempken *et al.*, 1992). For the Rpo-like sequences a domain that extended from the conserved Pol domain III to X was used (Kempken *et al.*, 1992).

Phylogenetic trees were derived from the mutiple sequence alignments using the tree drawing program in Mega ver 2.1 (www.megasoftware.net). The trees were derived using an unrooted neighbor-joining method (Saitou and Nei, 1987), and were constructed with the use of a Poisson distance matrix, one for Dpo and another for Rpo sequences. For the Dpo phylogeny (Figure 4.30), in which viral sequences (bacteriophage and viruses of eukaryotes) were the outgroup, the *L. perenne* sequences group with other plant linear plasmid sequences. Furthermore, within this clade, sequences from dicot species (*B. napus, B. vulgaris* and *D. carota)* [BS value 98%] form a discrete subgroup. The fungal sequences also clearly form a separate clade within which the basidomycetes plasmids (pEM, pFV2, pLMP1*)* have their own subgrouping [BS value 61%] relative to ascomycetes [BS values of 36%, 98%], except for the *Flammulina* pFV1 plasmid (Ryohsuke *et al.*, 2000) and the *Pleurotus* pLMF2 plasmid (Kim *et al.*, 2000) which may indicate that these two plasmids originated in a fungal species other than that in which they currently reside.

The Dpo of higher plant and fungal plamids are more closely related than are the Rpo, which accounts for the different groupings in Figure 4.31. In particular, the Dpo gene in the mt plasmids pCLK1, pBgh, and pAL2-1 forms a close association with those from higher plants whereas, in the Rpo tree (Figure 4.31), this grouping appears to be more closely associated with a different group

of mitochondrial linear plasmids [BS value of 49%]. However, higher plant-encoded Rpo form their own separate grouping with the *L. perenne* sequences forming a discrete subgroup [BS value 100%]. The plant and fungal sequences are well-separated from the nuclear-encoded, phage-like single subunit Rpos that are targeted to chloroplasts and/or mitochondria in higher plants. The *Neuospora* plasmids, *Kalilo* and *Maranhar* form their own separate branch. The Dpo and Rpo of the *Kalilo* plasmid is closely affiliated with several *Basidomycete*-like linear plasmids i.e. pFV1, pLMP2. The maize S1 (Dpo-containing) [BS value 82%] and S2 (Rpo-containing) [BS value 82%] plasmids of maize form their own branch and even though the S1 and S2 BS value is low, the association with other higher plant plasmids is consistent for both phylogenetic trees (Figure 4.30, 4.31).

4.4.7 Dpo and Rpo transcripts are detected in *L. perenne* lines containing the CMS cytoplasm

The *L. perenne atp9* gene shares 90% and 92% homology with the corresponding *atp9* genomic sequences of rice (Accession No. X16936) and wheat (Accession No. X15083), respectively. Upstream of *atp9* at nt position 104-349 is a putative promoter region sharing 99% sequence homology with the promoter region in the co-transcribed *atpA-atp9* gene region of wheat (Accession number X54387). Also present is the conserved element 5' TCGAAATAGCGTATAAA 3' found in several wheat mitochondrial promoter regions.

Northern blots of total RNA from the *L. perenne* CMS-St and Rst1 lines, probed with *orfC9* and *atp9* gave transcript sizes of 1.0 kb and 1.4 kb respectively as shown in Chapter 3. However hybridising of the above northern blot with a PCR amplified 2.9 kb LpDpo-specific probe and the diagnostic 4.5 kb *Bam*HI/*Hind*III (Kiang and Kavanagh, 1996; see Figure 4.13) utilised as a LpRpo-specific probe, failed to detect transcripts (data not shown). This suggests that LpRpo and LpDpo may not be transcribed or that transcripts are of low abundance. For this reason, it was decided to use the more sensitive RT-PCR approach to detect transcripts from CMSi.

As shown in Figure 4.32, RT-PCR products of the correct size for LpRpo, LpDpo and *orf3* were amplified from single stranded cDNA (+RT) for the *L. perenne* lines CMS-St and the restored line Rst1, both of which contain the CMS

cytoplasm. The PCR products DP1 (637 bp) and DP2 (684 bp) represent 43.9% of LpDpo; the 1,022 bp R1 and 843 bp R2 products represent 48.0% of LpRpo; and the 678 bp N1 product represent 61% of orf3. The negative controls (−) were RT-PCR reactions performed in the absence of reverse transcriptase (−RT) which gave no amplification product, in contrast to the positive genomic control (+g) as shown in Figure 4.32.

Northern blot data indicate that the chimeric *orfC9* gene is co-transcribed with the upstream *atp9* gene (Chapter 3). RT-PCR also detected co-transcripts: the DPN1 primer pair amplified a fragment of 1.474 kb (Figure 4.32) indicating the upstream LpDpo and downstream orf3 are transcribed as part of the same mRNA transcript. RT-PCR failed to detect a 2.2 kb co-transcript spanning the 3' termini of LpDpo and LpRpo, thereby failing to detect a transcript for *orf3* in this direction (data not shown).

All fragments of cDNA amplified by PCR (Figure 4.32) were sequenced to assess the occurrence of editing. However, no evidence for either C to T or the rarer T to C editing was found. This might indicate that integration of CMSi into the mitochondrial genome of *L. perenne* occurred relatively recently. Furthermore, no size difference was observed between PCR products from cDNA compared with those from mt genomic DNA, ruling out the presence of introns.

4.4.8 Expression of an LpDpo fusion protein in *E. coli*

In order to generate antibodies that could be used to determine if the LpDpo protein encoded by LpCMSi is expressed in CMS *L. perenne* lines, fragments of ORF1 were cloned into the *E. coli* recombinant protein expression vectors pTwin Impact (www.neb.com) and pET32a (www.novagen.com).

4.4.8.1 LpDpo fusion protein expression using pTwin-Impact

A 2.9 kb fragment of LpDpo, was PCR amplified from the pCMS60 clone using *Pfu* DNA polymerase and the primer pair pTwNcoIF and pTwBamH1R. The 2.9 kb fragment was cloned into the pTwin-Impact vector (New England Biolabs) at the *Bam*H1 and *Nco*I cloning sites. Recombinant protein expressed from this vector contains an N-terminal chitin binding domain (CBD) followed by a self-cleaving intein at the N-terminus of protein of interest, which increases the

molecular mass of the expected polypeptide by 26 kDa. Thus, the expected CBD-I-LpDpo polypeptide should have an unprocessed mass of 136 kDa. The pTwinLpDpo construct was confirmed by restriction digestion and sequencing (Figure 4.33). An *NcoI/Bam*HI double digest of the pTwin-Impact vector yields fragments of 829 bp and 6.54 kb. The latter fragment is the relevant vector fragment, while the 829 bp fragment is replaced by the gene of interest. Thus an *NcoI/Bam*HI of the correct clone contains two fragments of 6.54 and 2.9 kb (Figure 4.33). The *E. coli* expression host strain ER2566 was then transformed with the pTwinILpDpo construct.

Expression of the expected 136 kDa CBD-I-LpDpo fusion protein was found to occur optimally during 16 to 20 hr., incubated at a temperature of 18°C and an inducing concentration of 0.4 mM IPTG. Figure 4.34A shows fusion protein expression in the total protein fraction (whole cell lysate), which is just about visible on a Coomassie blue-stained gel. Sonication of the induced cells and recovery of insoluble and soluble protein fractions showed the 136 kDa protein to be found predominantly in the soluble fraction (Figure 4.34B). The anti-CBD antibody supplied with the expression vector was used to probe the soluble and total cell fraction by western blotting which confirmed the expression of the 136 kDa CBD-I-LpDpo fusion protein (Figure 4.34C). However, purification of the protein proved to be difficult because it would not bind with sufficient efficiency to the chitin purification matrix

4.4.8.2 LpDpo fusion protein expression using pET32a

Given the low yield of protein expression and difficulties encountered in its purification, it was decided to express a smaller LpDpo fragment in the vector pET32a (Novagen) which places a thioredoxin (Trx) domain, His and S-Tag at the N-terminus of the protein of interest. An 844 bp LpDpo fragment encoding part of the N-Terminal region and the conserved 5'-3' exonuclease domains EI, EII, EIII was PCR-amplified from pCMS60 using *Pfu* DNA polymerase and the primers TpF and EIIIR and cloned into *Bam*HI and *Hind*III cloning sites of pET32a to give pET32LpDpo as shown in Figure 4.35. The cloned LpDpo fragment is predicted to encode a 31.3 kDa polypeptide to which the Trx-His-S tag adds a further 18.7 kDa i.e. the expected fusion protein should have a mass of approximately 50 kDa

Optimal expression of pET32a-Trx-LpDpoTpEIII fusion protein was achieved in *E. coli* BL21(DE3) following induction of cultures with 1 mM IPTG at mid-exponential phase ($O.D._{600}$ = 0.5) and growth at 30°C for a period of 3hr. As shown in Figure 4.36 induced but not uninduced cultures expressed a protein with the expected mass (50 kDa) at high level. Disruption of the cells by sonication revealed the expressed protein to be located exclusively in the insoluble fraction (lane 8, Figure 4.36A). The insoluble fusion protein was purified by means of electroelution (Biorad) into a 1X SDS elution/running buffer, and was then dialysed into PBS buffer (pH 7.1) (see Chapter 2 section 2.2.10.3).

Figure 4.36B shows the samples of purified fusion protein that were used to raise polyclonal antibodies in rabbits. The first injection (Freunds complete adjuvant) amounted to approximately 150 µg fusion protein followed by a two additional booster injections of 100 µg in Freunds incomplete adjuvant). The resulting anti-Trx-His-S-LpDpoTpEIII antibodies were initially tested at a 1/15,000 dilution in a western blot analysis of induced and uninduced *E. coli* cell extracts that contained the pET32-Trx-LpDpo construct (Figure 4.37). As shown in Figure 4.37B the antibodies specifically detected a 50 kDa protein in induced but not uninduced extracts.

Unfortunately, although the anti-Trx-LpDpo antibodies worked well in the detection of the Trx-LpDpo fusion protein in *E. coli* cell extracts, it did not detect LpDpo protein in mitochondria-enriched protein extracts from fertile or CMS *L. perenne*. This indicates that the LpDpo protein may not be produced in CMS mitochondria, or that it is produced at very low levels, or that the specific affinity of the anti-Trx-LpDpo antibodies for their target antigen is too low (Figure 4.38).

4.5 Discussion

LpCMSi represents the first documented case in a monocot species of an integrated linear plasmid-like element containing both an intact *Dpo* and an intact *Rpo* orf. Remarkably, despite extensive sequence divergence, the arrangement of the orfs in LpCMSi is similar to that in the *B. napus* 11.6 kb extrachromosomal linear plasmid, in which the polymerase orfs are located on opposite strands and transcribed towards each other. This inverted tail-to-tail *Dpo-Rpo* organization is also found in linear mitochondrial plasmids of the filamentous fungi *Claviceps purpurea* (e.g. the p*CLK1* plasmid-Oeser and Tudzynski, 1989) and *Neurospora* (e.g. the *Maranhar* and *Kalilo* plasmids-Court *et al.*, 1992; Chan *et al.*, 1991). In contrast, the *Rpo* and *Dpo* orfs of the extrachromosomal linear plasmids of *B. vulgaris* subsp. *maritima* (Saumitou-Laprade *et al.*, 1989) and *D. carota* (Robison and Wolyn, 2005) are located on the same DNA strand in a head-to-tail arrangement.

Investigations of gene expression in linear plasmids has mainly focussed on the detection and quantification of orf transcripts, so that only limited data concerning transcription initiation and the identification of promoter sequences is available. Transcripts have been shown to initiate in the TIRs of the fungal plasmids *pCLK1*, *Kalilio*, *Maranhar* and *Bgh* (Gesser and Tudzynski, 1992; Vickery and Griffiths, 1993; Giese *et al.*, 2003) and a putative yeast-like ARS consensus sequence (-ATATAAGTA-) that may function as a promoter element was localized within the TIR region of the *Kalilo* (Vickery and Griffiths, 1993) and *Bgh* plasmids (Giese *et al.*, 2003). In the case of LpCMSi, the highly conserved promoter of the *atp9* gene (Covello and Gray, 1991) located immediately upstream of LpCMSi, most likely drives transcription of *orfC9* which is co-transcribed with *atp9* (Chapter 3) at a level that is detectable by northern blot hybridisation. However, given the very low level of transcripts for the downstream orf, which were not detectable by northern blotting, it is unlikely that transcription is under the control of the *atp9* promoter. Potential promoter sequences driving low level expression of the Rpo transcript and the *Dpo/orf3* co-transcripts (which were detectable only by RT-PCR) were not identified.

With the exception of *L. perenne* (this study), *D. carota* (Robison and Wolyn, 2005), and certain *Z. mays* accessions (Garcia *et al.*, 1988), the majority

of Dpo- and Rpo-like sequences found in plant mtDNA consist of orf fragments, the products of putative deletion or insertion events that occurred following plasmid integration. The completely sequenced mitochondrial genomes of *B. vulgaris* (Kubo *et al.*, 2000; Satoh *et al.*, 2004), *O. sativa* (Notsu *et al.*, 2002) and *Z. mays* (Clifton *et al.*, 2004) contain both Dpo and Rpo orf fragments. However, the mitochondrial genome of *T. aestivum* (Ogihara *et al.*, 2005) contains only Rpo-like sequences while that of *A. thaliana* (Unseld *et al.*, 1997), *N. tabacum* (Sugiyama *et al.*, 2005) and *Marchantia* (Oda *et al.*, 2002) contain only Dpo-like sequences. Although not yet fully sequenced, *S. cereale* mtDNA appears to lack Rpo-like sequences but contains a truncated Dpo orf encompassing the Pol IIa to Pol III domains (Dohmen and Tudzynski, 1994).

Surprisingly, neither the presence nor the absence of Dpo and or Rpo sequences in plant mtDNA correlates with the presence or absence of the corresponding homologous plasmids. Linear mitochondrial plasmids have not, for example, been reported in *L. perenne*, *A. thaliana*, *T. aestivum*, *O. sativa*, *S. cereale*, *N. tabacum*, or *Marchantia,* and remarkably, despite the common occurrence of a high copy number linear plasmid in *Brassica* mitochondria, the sequenced *B. napus* mitochondrial genome contains neither Rpo- nor Dpo-like orf fragments (Handa, 2002). Similarly, the *D. carota* mitotype FG21 contains episomal but not integrated copies of a 9.2 kb linear plasmid, while the SW3 mitotype contains integrated but apparently not episomal copies of the same plasmid (Robison and Wolyn, 2005). In sugar beet (*B. vulgaris*), the completely sequenced mitochondrial genome of a fertile line (369 kb) (Kubo *et al.*, 2000) and a CMS line (501 kb) (Satoh *et al.*, 2004) contained an identical 6,253 bp linear plasmid-like region encoding disrupted Dpo and Rpo orfs. However, the nucleotide sequence of this region was only 43% homologous to that of the episomal 10.4 kb linear plasmid identified in *B. vulgaris* subsp. *maritima* (Acc No.Y10854), from which it was most likely derived. Similarly, the completely sequenced *Z. mays* (NB) mt genome contains 2 integrated Dpo sequences, *orf911* and *orf734*, remnants of independent S1 linear plasmid integration events. Remarkably *orf911* which is almost full-length, shares only 63% amino acid identify with the Dpo encoded by extrachromosomal copies of the S1 plasmid, while *orf734*, which lacks the N-terminal 200 amino acids shares 99.5% amino acid identity with S1 (data not shown). These observations suggest that in contrast

to the mitochondria of filamentous fungi, plant mitochondria possess efficient mechanisms that inhibit the integration of linear plasmids and disable or eliminate occasionally integrated copies by mutation.

Neither Southern hybridisation with the Dpo and Rpo probe, nor PCR using primers that amplify the junction fragments (JL and JR), detected LpCMSi-like sequences in the mtDNA of fertile *L. perenne* lines. However, PCR based on primers annealing within the LpRpo and LpDpo reading frames revealed the presence of LpCMSi-like sequences in FNz and FIp4. Moreover, amplification with primers annealing to the 3'ends of the Rpo and Dpo orfs yielded a 2.2 kb product containing the *orf3* region, indicating that the Rpo and Dpo reading frames in FNz and FIp4 are contained on the same replicon (data not shown). Taken together, these data suggest that FNz and FIp4 mitochondria contain LpCMSi-like elements that are maintained either as very low-copy number episomal replicons, as has been reported for certain linear mitochondrial plasmids in fungi (Robison and Horgen, 1998), or as integrated elements on low-copy number (i.e. substoichiometric) subgenomic mtDNAs (Small *et al.*, 1987). It may be significant that the CMS-St line was generated during a hybrid ryegrass breeding programme from an interspecies cross between *F. pratensis* (as the female parent) and *L. perenne* (Connolly and Wright-Turner, 1984), which unexpectedly resulted in inheritance of the paternal mitochondrial genome in the hybrid (Kiang *et al.*, 1994). Wide crosses are known to promote illegitimate and asymmetric recombination events that can lead to rearrangements in mtDNA and dramatic changes in the stoichiometry of subgenomic molecules, so-called substoichiometric shifting (SSS) (Abdelnoor *et al.*, 2003). Thus, in the case of LpCMSi, the *Festuca* x *Lolium* cross may have promoted the original linear plasmid integration event or, alternatively through SSS, dramatically increased the copy number of a subgenomic replicon in which LpCMSi already resided.

Although our phylogenetic analyses do not appear to support a fungal origin for the LpCMSi sequences, horizontal gene transfer between *L. perenne* (or *F. pratensis*) and a fungus remains a possibility. Fungi of the genera *Epichloe*, *Neotyphodium* and *Acremonium* are mutualistic symbionts (Latch *et al.*, 1984; Byrd *et al.*, 1990) of grasses (especially of *Festuca* and *Lolium* spp.), and most notably with linear plasmids being contained in *E. typhina* (Mogen *et al.*, 1991). During the asexual phase of their life-cycle these fungal symbionts grow in close

association with developing floral organs including ovules, a strategy that ensures their transmission via seeds (Schardl *et al.*, 2004). It is not inconceivable that such a close association between fungal mycelia and plant reproductive organs might also facilitate rare horizontal transmission of genetic information between the symbiont and its host. Horizontal transfer between fungi and higher plants is not without precedent. The best documented case involves a mitochondrial homing group I intron of fungal origin which invaded the *coxI* gene of flowering plant mitochondria on more than 1,000 occasions during angiosperm evolution (Vaughn *et al.*, 1995; Cho *et al.*, 1998). The LpCMSi orfs, like those of other linear plasmids, exhibit a significantly lower % GC value, closer to that of fungal mtDNA than the average for plant mtDNA. Nevertheless, LpCMSi sequences grouped with plant and not fungal plasmids. This discrete grouping might, however, simply reflect differences in mutation rates between plant (Cho *et al.*, 2004; Mower *et al.*, 2007) and fungal mitochondria following an evolutionarily distant horizontal transfer event, and does not rule out the possibility that fungal plasmids may be capable of replicating in plant mitochondria or that horizontal transfer might have played a role in the acquisition of LpCMSi by *L. perenne*. An extensive PCR-based survey of *Lolium* and *Festuca* species and their fungal symbionts may shed more light on the origins of LpCMSi.

Finally, we sought to express LpDpo in *E. coli* in order to obtain soluble recombinant protein that could be used to (a) raise antibodies for detection of LpDpo by western blot analysis and by immunohistochemical methods and (b) explore its biochemical properties i.e. DNA binding, exonuclease activity and polymerase activity. Unfortunately, as for many plant proteins expressed in *E. coli*, and despite using different expression strategies, the expressed proteins were either insoluble or could not be purified to homogeneity. Moreover, the antibodies raised against insoluble recombinant LpDpo when used in western blot analyses did not provide evidence of LpDpo expression in mitochondria of the CMS line. It might nevertheless be worth pursuing these goals using a different strategy in the future.

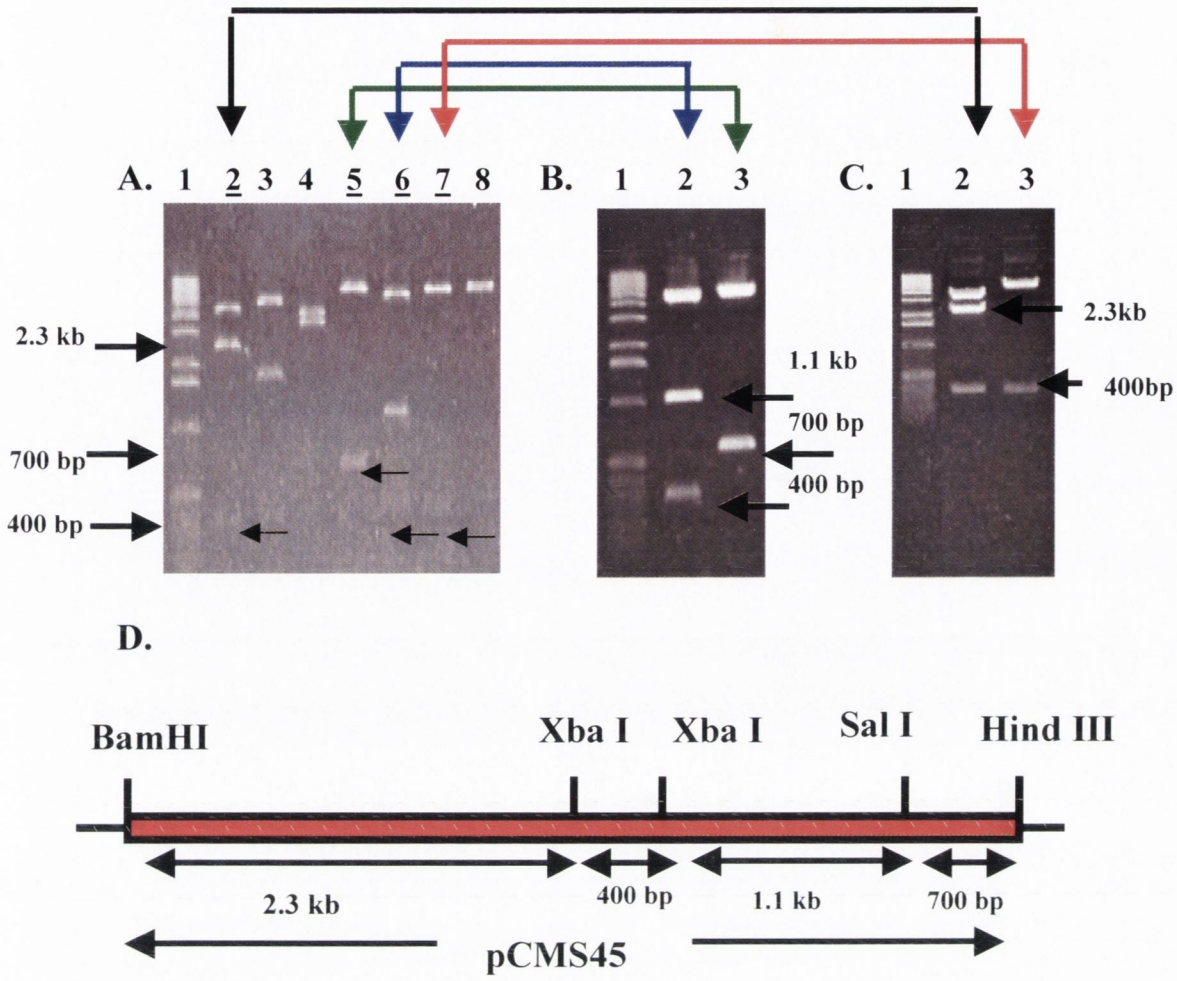


Figure 4.1 Restriction map of the pCMS45 portion of the CMS associated mtDNA region in *L. perenne*.

A. Lanes 2 to 6 are double digests of pCMS45 namely; *Bam*HI/*Xba*I, *Hind*III/*Xba*I, *Bam*HI/*Sal*I, *Hind*III/*Sal*I and *XbaI/SalI* respectively. Lane 7 and 8 are single digests of *Xba*I and *Sal*I respectively. Lane 1 is the 1 kb marker.

B. Lanes 2 and 3 are double digests of *XbaI/Sal*I and *Hind*III/*Sal*I respectively, thereby confirming the presence of a 1,100 bp, 700 bp, 400 bp fragment size that is not clearly visible in Figure 4.1A. Lane 1 is the 1 kb marker.

C. Lane 2 is a double digest of *Bam*HI/*Xba*I that also confirms the presence of a 400 bp size fragment that is not visible in lane 2 of Figure 4.1A. Lane 3 is a single *Xba*I digest and once again reconfirms the 400 bp fragment that is not as visible in lane 7 of of Figure 4.1A. Lane 1 is the 1 kb marker.

D. Structure of the 4.5 kb fragment in the pCMS45 clone showing the relative positions of the restriction sites.

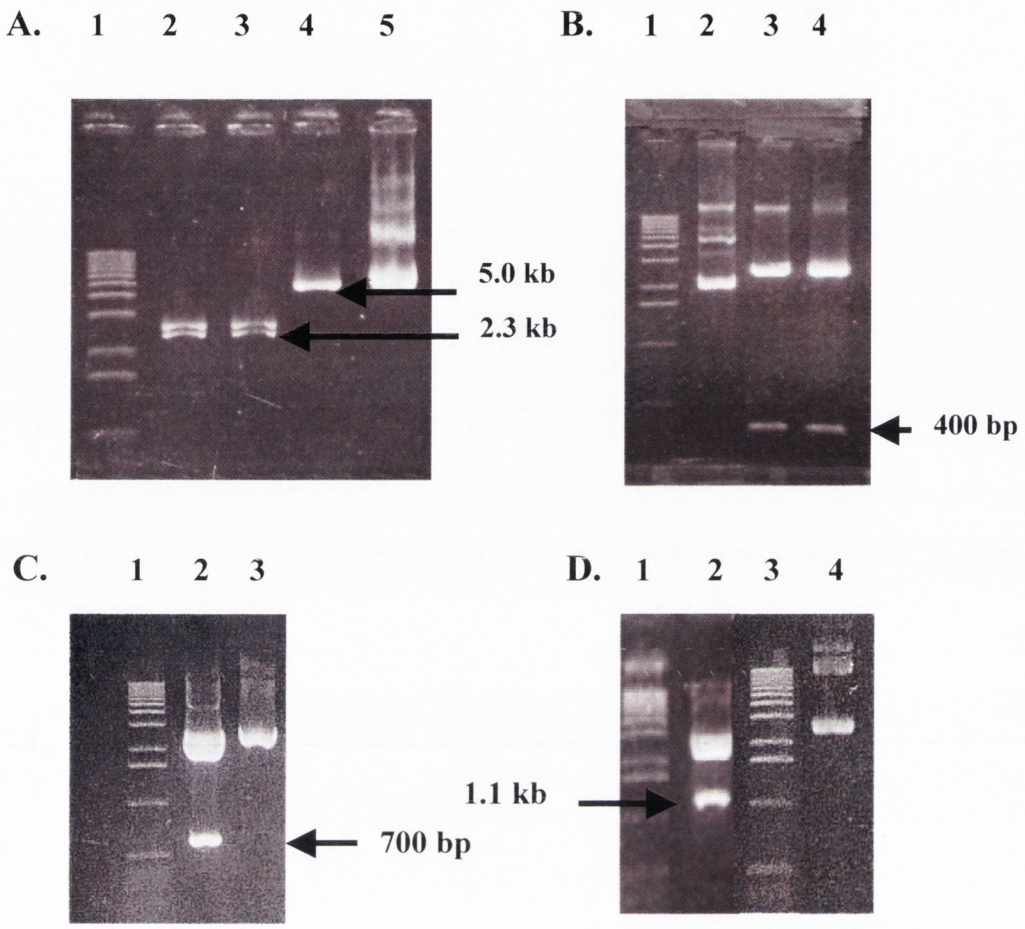


Figure 4.2 Restriction digests confirming the cloning of the subclones pCMS23, pCMS11, pCMS7 and pCMS4 in pUC19.

A Lanes 2 and 3 is a double *Bam*HI/*Xba*I digests of pCMS23 subclone, lanes 4 and 5 is a single *Xba*I digest of pCMS23 and undigested pCMS23, lane 1 is the 1kb DNA marker.

B Lanes 3 and 4 is a single *Xba*I digest of pCMS4, lane 2 is undigested pCMS4 and lane 1 is the 1 kb DNA ladder.

C Lanes 2 and 3 refer to a double *Sal*I/*Hind*III digest of pCMS7 and undigested pCMS7 respectively, and lane 1 is the 1 kb DNA ladder.

D Lanes 2 and 4 represents a double *Sal*I/*Xba*I digest of pCMS11 and undigested pCMS11 respectively. Lane 1 and 3 corresponds to the 1 kb DNA ladder.

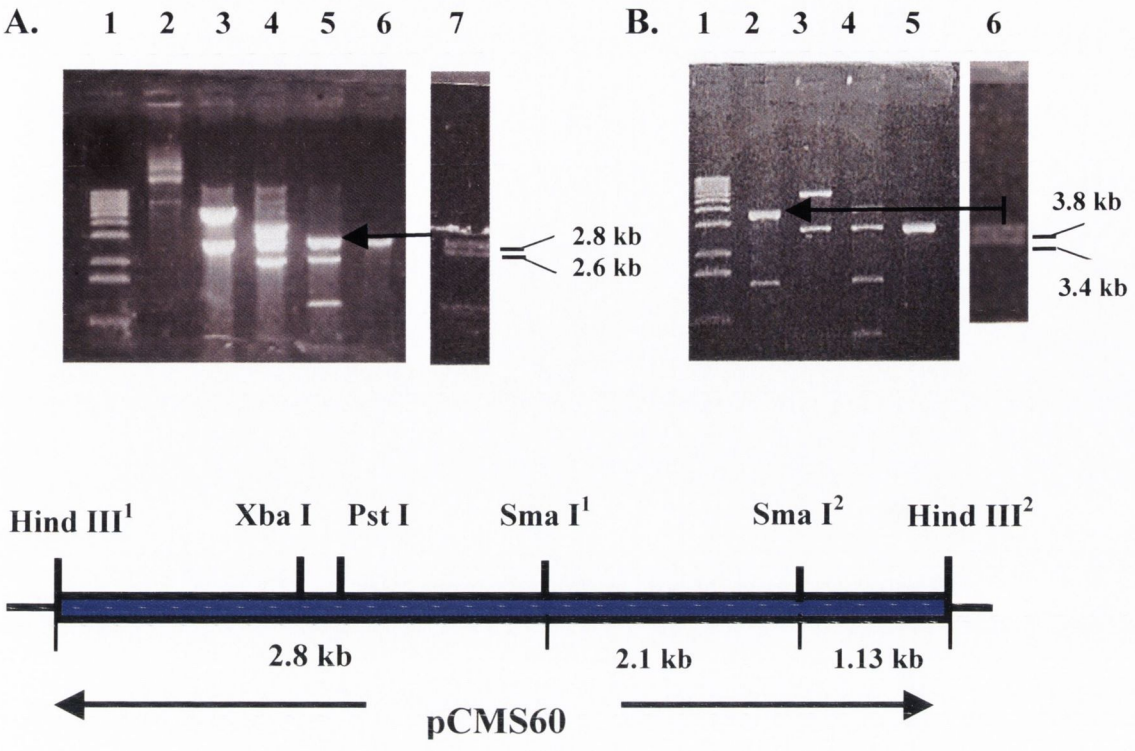


Figure 4.3 Restriction digestion analysis of pCMS60 clone.

A. Lane 2 is uncut pCMS60, lanes 3, 4 and 5 are single *Hind*III, *Sma*I digests and a double *Hind*III/*Sma*I digest, respectively. Lanes 6 is linearised *Hind*III/*Sma*I digested pUC19 plasmid. Lane 7 is lane 5 in that it has been run for a longer period, thereby illustrating further separation of bands for 2.6 kb (vector) and 2.8 kb (fragment). Lane 1 is the 1 kb marker.

B. Lanes 2 to 4 are single digests of pCMS60 by *Pst*I and *Hind*III and a double digest of *PstI*/*Hind*III respectively. Lanes 5 is linearised *PstI*/*Hind*III pUC19 plasmid. Lane 6 is lane 2 that was run for a longer period, showing a gel separation of two bands with sizes of 3.8 kb and 3.4 kb respectively. Lane 1 is the 1 kb marker.

C. Structure of the relevant restriction digestion sites in the 6.0 kb fragment.

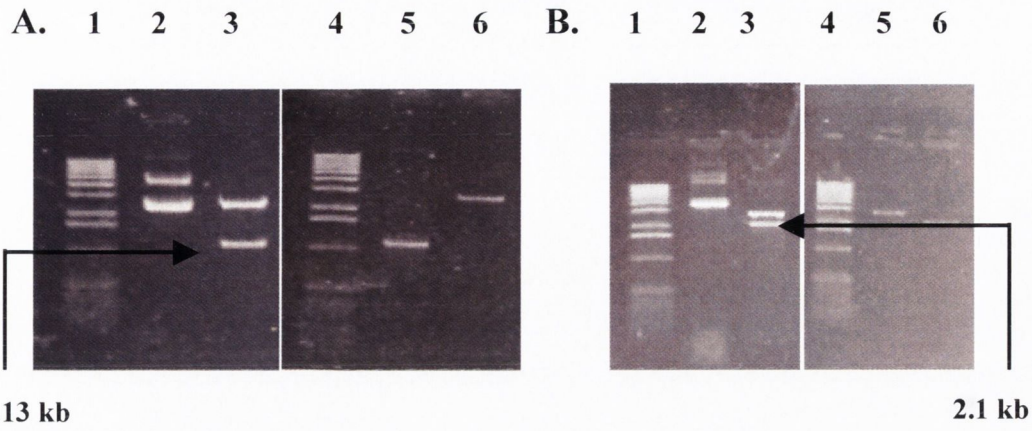


Figure 4.4 Restriction digest confirming the identity of the subclones pCMS10 and pCMS21 generated from pCMS60.

A. Lanes 2 and 3 are uncut pCMS10 and double *Hind*III/*Sma*I digest of pCMS10 respectively. Lanes 5 and 6 is the 1.136 kb fragment and linearised *Hind*III/*Sma*I digested pUC19 vector respectively. Lane 1 and lane 4 is the 1kb DNA marker.
B Lanes 2 and 3 is uncut pCMS21 and a single *SmaI* digest of pCMS21 respectively. Lanes 5 and 6 is the linearised *Sma*I digested pUC19 vector and the 2.1 kb fragment respectively. Lane 1 and 4 is the 1 kb DNA marker.

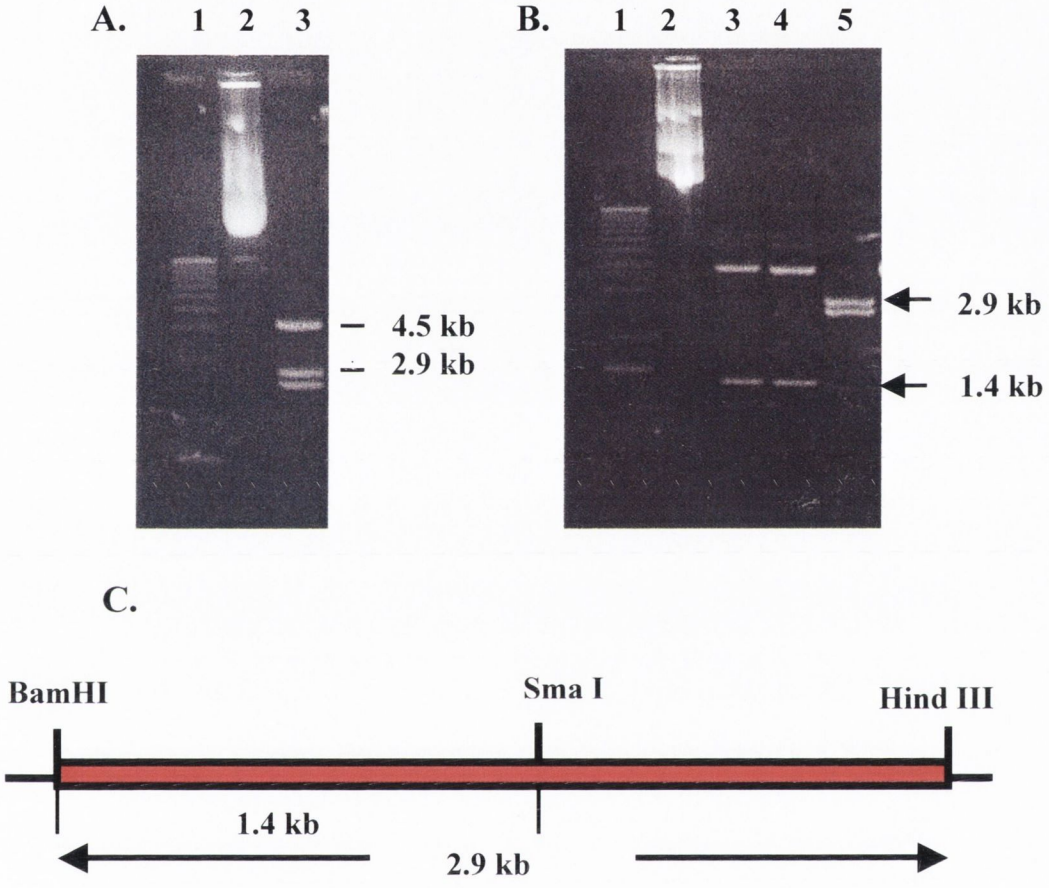


Figure 4.5 Restriction digest confirming the identity of pCMS29 subclone.

A. Lane 2 is uncut pCMS74 and lane 3 is a double *Bam*HI/*Hind*III digest of pCMS74. Lane 1 is the 1 kb DNA marker. **B.** Lane 2 is uncut pCMS29, lanes 3 and 4 are double *Bam*HI/*Sma*I digests of pCMS29 respectively. Lane 5 is a B*am*HI/*Hind*III is a double *Bam*HI/*Hind*III digest of pCMS29. Lane 1 is the 1 kb DNA molecular marker. **C.** Structure of digestion pattern for 2.9 kb size fragment.

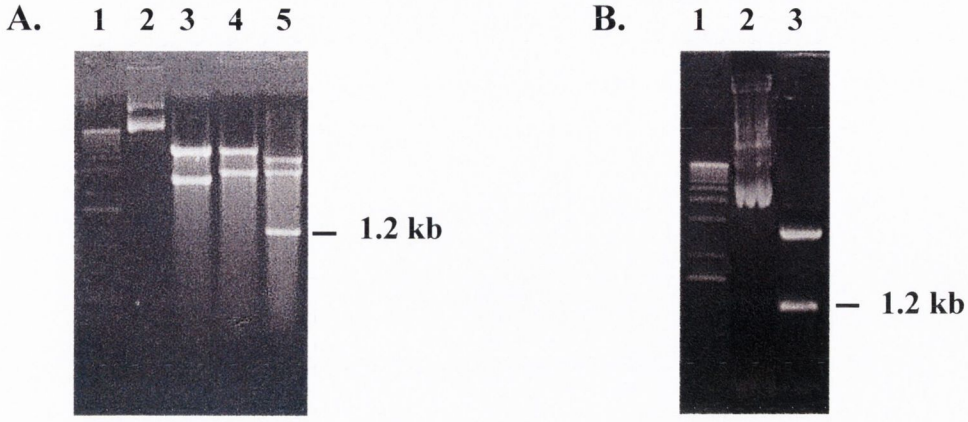


Figure 4.6 Restriction digest of the fertile pLMF56 and confirmation of the pLMF12 subclone.

A. Lanes 2, 3, 4, and 5 is undigested pLMF56, *Hind*III digested pLMF56, *Eco*RI digested pLMF56, *Eco*RI/*Sma*I digested pLMF56 respectively. Lane 1 is the 1 kb DNA ladder. **B.** Lanes 2 and 3 are undigested pLMF12 and digested *Eco*RI/*Sma*I pLMF12 respectively. Lane 1 is the 1 kb DNA ladder.

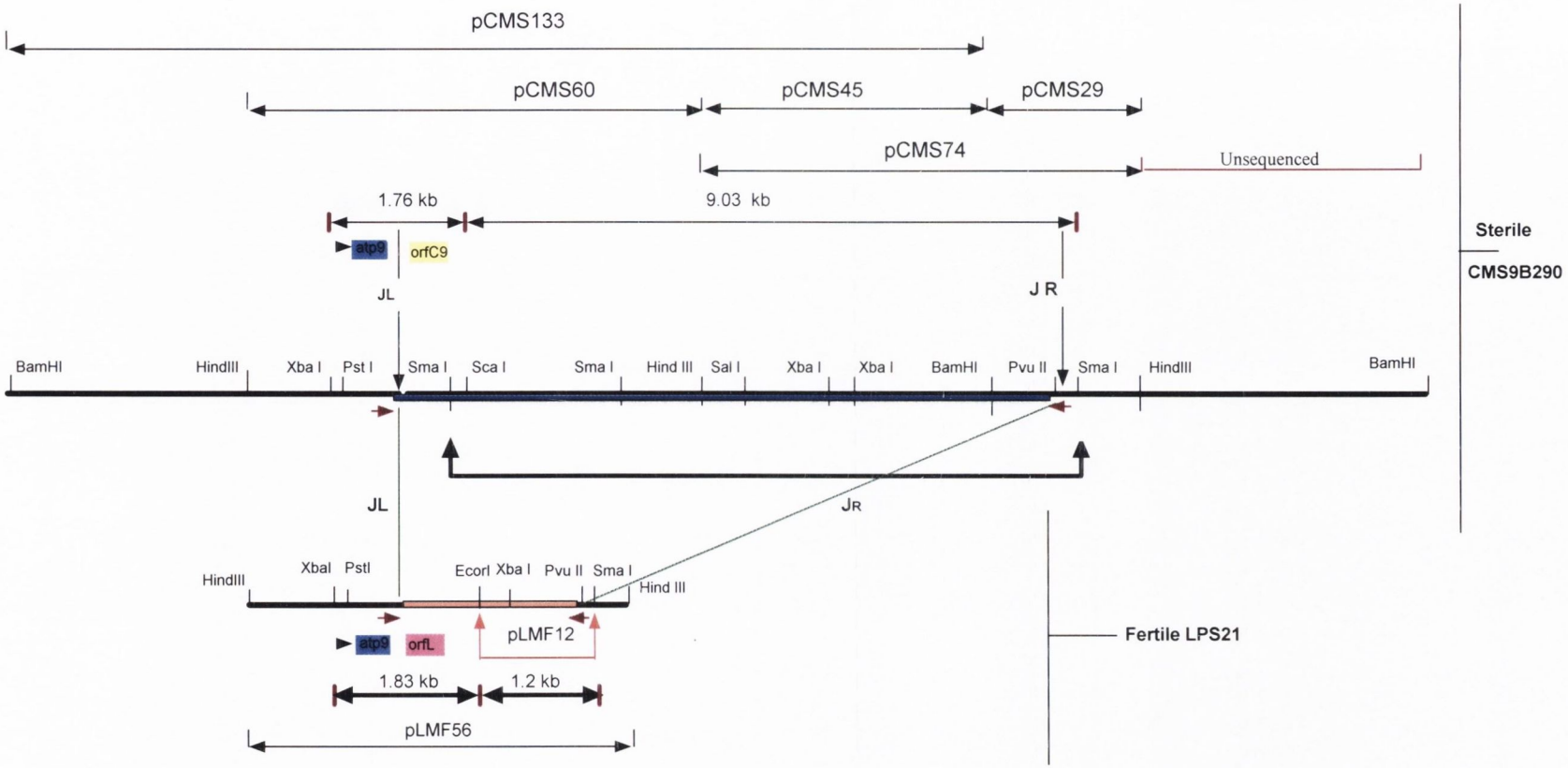


Figure 4.7 Restriction of the CMS associated region in CMS9B290 mtDNA and in the fertile line LPS21 mtDNA in *Lolium perenne*

The non-homologous regions in the cms and fertile line are shown by the blue and orange boxe lines respectively and the homologous region of both *L.perenne* lines is represented by the black box line. The clones pCMS133, pCMS60, pCMS45, pCMS 29, pCMS74, are indicated for the CMS associated non-homologous region and the 9.03 kb region for sequencing is highlighted. The subclone pLMF1.2 for the fertile non-homologous region is also indicated. The pale blue box is the mitochondrial atp9 gene, the yellow and purple boxes are representative of the chimeric genes orfC9 and orfL mitochondrial genome of the CMS and fertile lines respectively. The junction sites **JL** and **JR** are indicated by the arrow symbol → and ← respectively. The atp9 promoter for wheat and rice (Gray and Covello, 1991) is represented by the symbol ▶ .

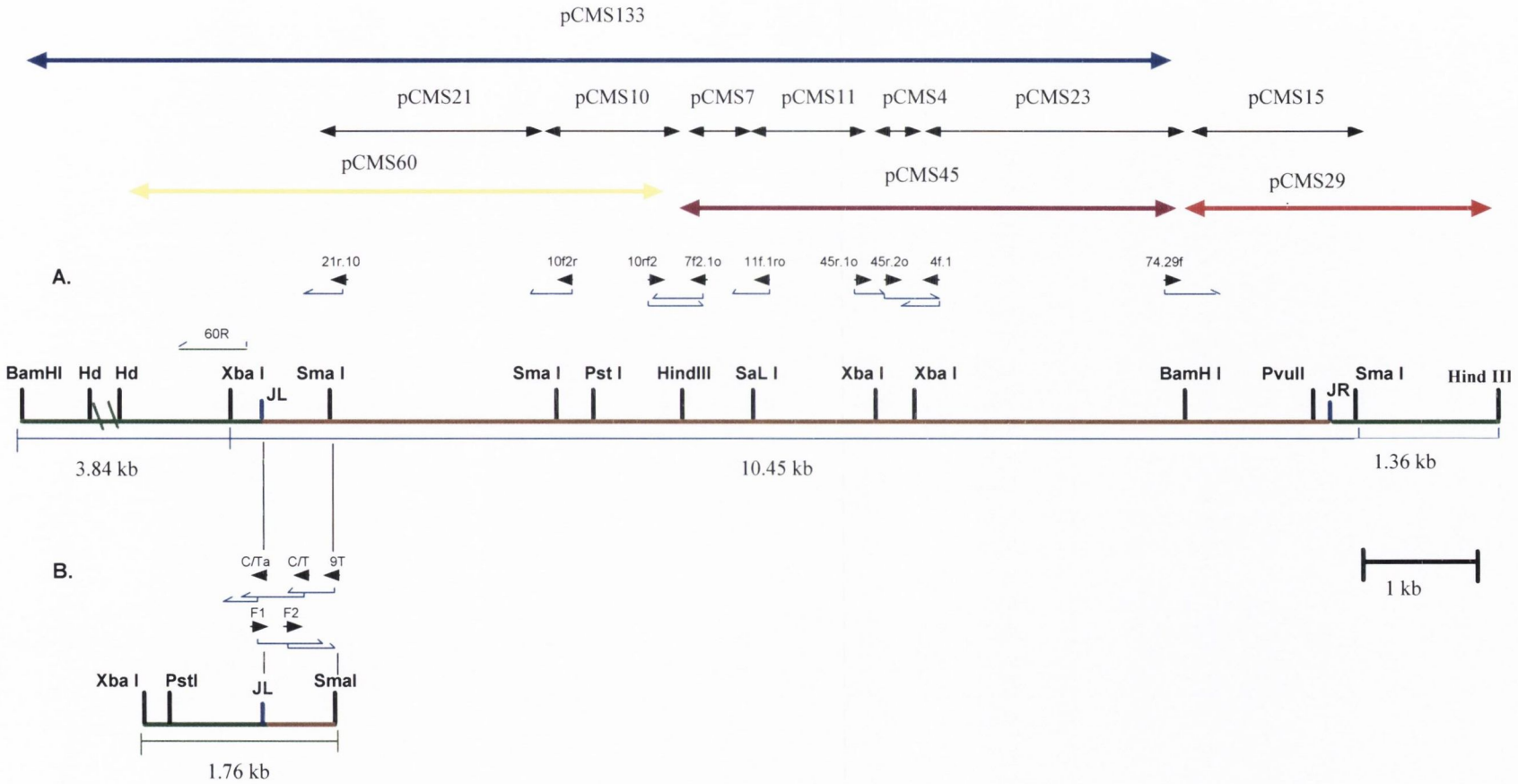


Figure 4.8 Restriction and sequencing map of homologous and non-homologous regions in sterile line CMS9B290

The non-homologous region associated with sterility in the CMS line-CMS9B290 is indicated by the line ▬ whereas the homologous sequence region is shown by the line ▬. **A.** The primers used to confirm the overlaps of the subclones pCMS23, pCMS21, pCMS4, pCMS7, pCMS11, pCMS10 are represented by the arrow symbol ➔. **B.** Shows the primers used (courtesy of Kiang and Kavanagh, 1996) to confirm the previously sequenced 1.76 kb region in fertile maintainer line LPS21.

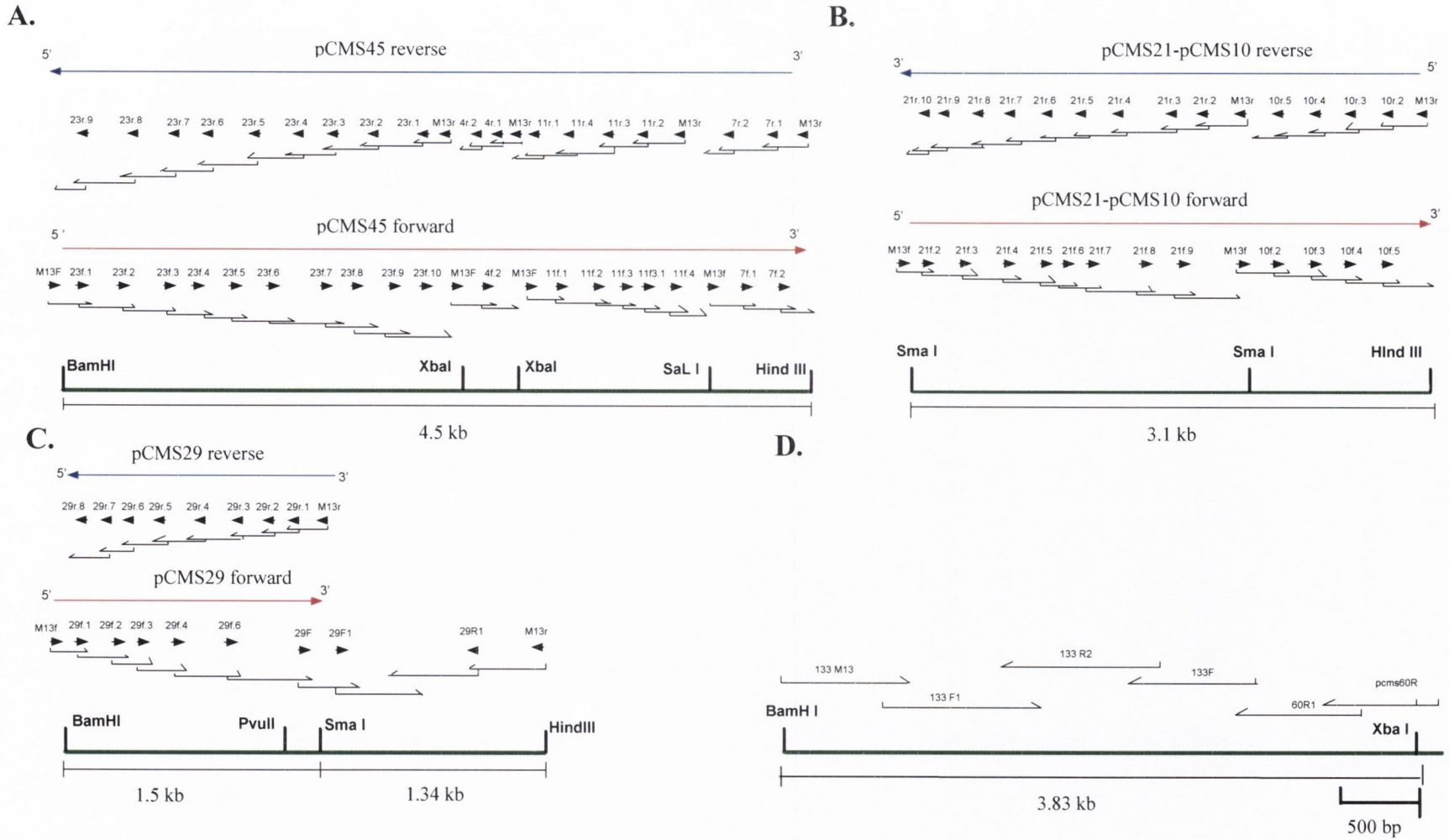


Figure 4.9 Represents a series of primers used in the sequencing of a 14.23 kb region in mitochondria of the line CMS9B290

A. Primers used in the sequencing of 4.5 kb region for clone pCMS45 are represented by the arrow symbol ➔ . **B.** Primers used in the sequencing of a 3.1 kb region for the clones pCMS21 and pCMS10. **C.** Primers used in the sequencing of 2.84 kb region for the clone pCMS29. **D.** Primers used in the sequencing of 3.83 kb region for the clones pCMS133.

A.

BamHI

```
   1 GGATCCTGTC TAAATCCATC CCCGATTCAC TAATTTCTGG CGCATAAAAA

  51 CAGAGTAAGA TTCTCCACCT GCAAAAGCTC TGCCACGATT CCTTGTCTAA

 101 TAATAAAGGC ATTGGGATGA AGCGGTAAGT GGCACACATG ACCTACCATC

 151 CGCGGGAATC TACTAACCAC AAGGGCGTTA AGAGAACTCT TACTGGACTG

 201 GAAACGGCTA GTGCACGTCG AATTTGATAA AATGTTACCT GAGAGTGCTT

 251 ATGGCTCACC TTCCTGCTTC CTATATTTTC CCGGATTCAA CTAATAAGAA

 301 GCGCCCTCTT GCGAGAGAAA GTCGCCCATC ATATCTATAG AATAAGACCC

 351 CCAATCCTGC GGTGCCATCC ACGTCGCGCG CTTGATTCCT TTCTGTTTAA

 401 AACCATATCC TTTCTTTGCC TATTAGACTA CCAATTCTAT CTGTGATTTC

 451 GCTCTGGTCA CTAGTATTTG TATTATCGAA TATCTCCTTT TCACTACTAG

 501 TGTACGATGA GATTCTATGT GAGTTGTATG ATCAATGAAG TAATTGTATT

 551 CTGATTGCAT TTCTCGGATG AAATATTTAT CAAAGTCAAG AAGGAAGTTT

 601 CTAATATTAT TCATATTATA GCTATCTTTA GTCCCTATGT AAAAATAAGT

 651 AGAATATTGT ACTAGTAAAA TGAAATTGAG AATCATTATT TCTTTTTTCT

 701 GTATCTTTTT TGTTTCGGGA TTATGTCGAA TGATGTCAAT ACCCGTGTTG

 751 TGGCGTAAGC CACCTCTAGC GTTGTTGCTG GCTAGAGAAA GACACCCCTT

 801 TATGTATATA GCCAGTATTT AACCAGACCT ATAAATGATG GCCGTTGATC
                HindIII
 851 GGTCCCTCTT TTTCAAGCTT ATCTTTATAG ACGGCTGGGT GTGGGGGGTA

 901 CCTTTCTTTA GAACAAGTAT TCGTGATACC GGTTTCCCTC CCTCGGTCTA

 951 CAGTCAACTA CTGGGCTGCG GATTCAGGGA CAAGCAAACT GTGATGACAG

1001 TACCTGGAAC CCGAAAAGGG TGACCCCCTT AGGGTCAAGT AGTTCTCTTG

1051 TCTGATAAGA TCTATGGGTT TCATAGGATC ATTAGACGCC CGAGCGAGTA

1101 GGCATCGCAT CTTATATCTC CTGTATGATA GATAGTATAT AAGAGTTATT

1151 GAAACTAACA TACTATATAG CGTCTTTCAA GTATACTATC TCCCCACTAA

1201 CGACTATCTT TTTCAAGCGT AGAGCGAAGT CGATGAATGA TATACTTCTT

1251 TTAGCTGAAT CATTGACTCT GATCGGAACC TGGCGACCCC GAAGGGAGAA

1301 CAAGTTCAAA GAAAGAGATG TCCGAATGGT TAATAGCACG AGCTCTGCCC
```

1351 CATCTATTGA ATCATTCAAA CTGGAGAGTC CAGTATATTA AACATAGGGT

1401 ATAAGTATCA TAACCTATAC TCTAGTAAGA AATATTTACC ATGAAGACTT

1451 TCAAAGCGGG GAATAATCTT CATTCCTAAG ACAAGGGCCC CTCCGCTTTT

1501 TGGGGCATGG GGAAAGCATT TAAGAAGCAG TATGTGACAG GATATAACCA

1551 TTTACGCTGA GGACAATGAG ACTGACTTAG GCCTGCGAAC CTACCCCCAA

1601 GACTTTCATG TCTAACGGAA AGTGAAAGTG CTGAACGCAT TGCATTACTA

1651 GCTTTACGTT ATTCTCCTTA TATGCTAATT GTGTGAGAAT TAGCAGCTTA

1701 TAAGCATGTA CGCGATATCT GATCAATCGC TCCTGCCGGT TGTACAGATT

1751 CTATCACTGG CTTTAGGTTA TTCCTTCTCG GACCTGACGT CTCCATCACT

1801 ACTAATAAGG CCACTTTCTT TTGTATCGGG CCTTCAAAAC GCAACGCAGT

1851 AGAAACAGAT AGCTGAGTAG GGAGTGCGGC CGAAGCATCC ATAGCTTAGC

1901 GATTATGTTC TTATAGATAA GGATTAGACG GATGTAGACA ACCAAGAAAT

1951 CTAAGGTGAA CCAATAGAAA TAGAGTCTGG TGGAAGCGGT CATCCCGTGG

2001 GGAGAGACTC AAGAATATGA TAATATGATG ACTCCGCTGC GCTGCTCTGC

2051 TGACTAGCCG TTTCCTCTTC TAAGTAAATA ACCTGCTTCA GTTCATGCTC

2101 CTTGCTGGCA TAACCATATT AAGCTGCATT GGATAAACAA TCGAAACTCC

2151 GTAATTGTTA ATAGTAGGGT AGCGAAGTAG CCGGGGTTGC ACGCGTCATC

2201 AGAATTCTAG TATAAGGAAG CGACGGGAAG GGAAAGGTGA TTACAAAATA

2251 CGGATACGTA AGGGTTGTGT TTATTGACTA TCACAGGAAA ATGGTAATAA

2301 TTTATCACAC ATGGTATCAG AGCTGATACC TTCACCCAAA CCGTAGACGC

2351 CGGCGGTGAA GAAGGAGGTG AGGGCGCCTG CTTTGATTAG GCTTACTGTA

2401 AGTTTGTTTG AAGCTCCGGA GTCTAGTCTA CCAATGGGTG CTATGGTGTG

2451 TCCTCCCCTC TAAGGTCTCG CTTGGATAAT CGACTCTTAC GACCCACACA

2501 GGAGAGGGTT TTTCCCAACC GTTCATGCCT CTCGGCTATT AAATAAAAGA

2551 GTGGGCTCTT TGAAGCTAGG TATTAAGCAC AGGAAGGTAT GGGAATATAG

2601 CGAGCACTTT CTTAAACTTA AACCGAGATG TGAGACTACT AAAGAGGCAG

HindIII

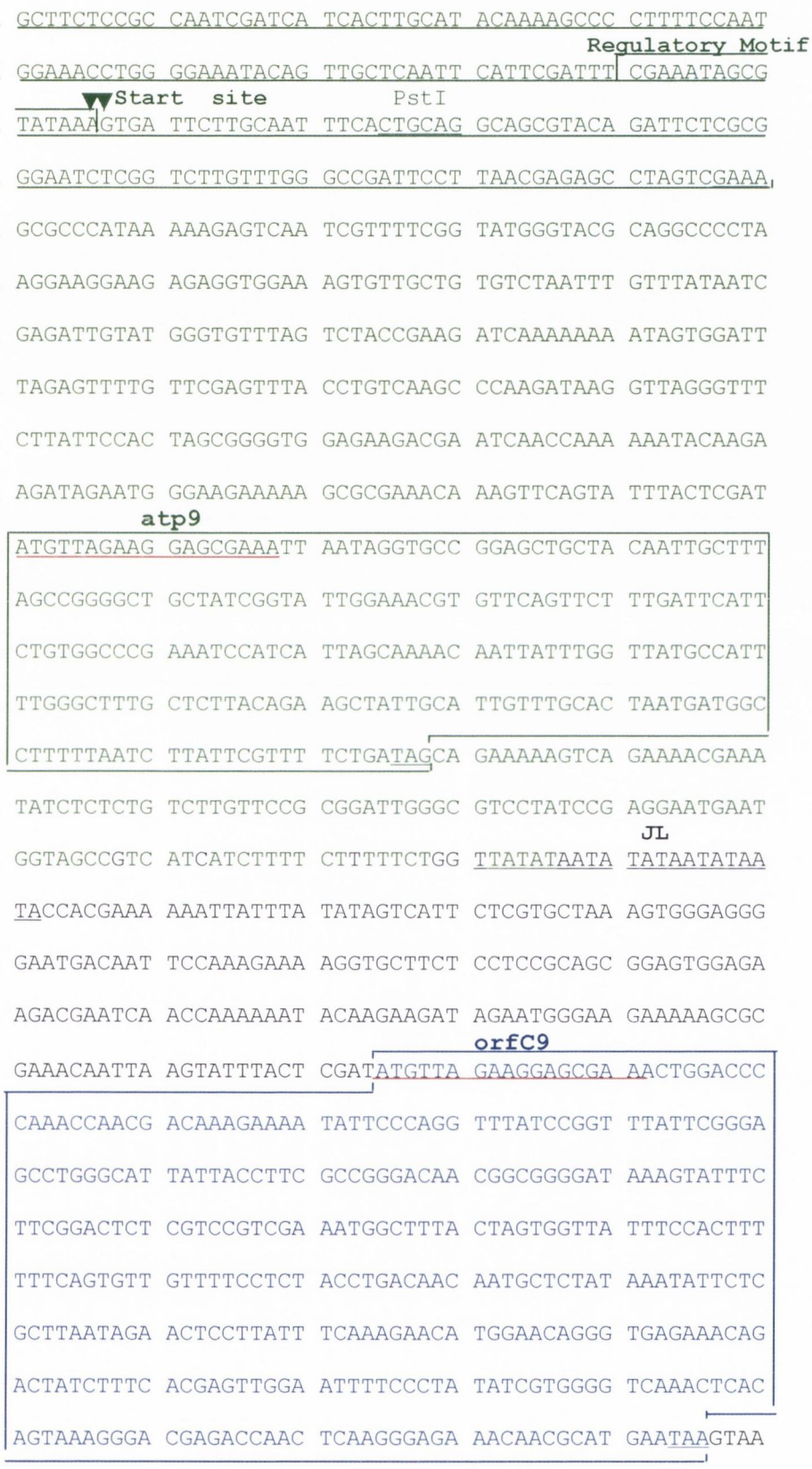
2651 GTAAAGTGGG AAGTGGGATT CATGATGCAA GCTTATGAAG AAAAAAAGAG

2701 ATCTCATTAA TGCGTAAAAA AAAAGAATAG CAATCCCTCG CCATCTGAGC

```
2751 CAGAACGAAT CCCAGTAAAA TTGCTAAATG CTTTTGGCAT TTTTTTCCCT
2801 CGCATTTCCC TAATTCTTTT ATTAATATGG GAATGGGAGC AAATGAGGAG
2851 GCATCGAGTC GAAGAAATGA ATTGCTCTTT GGCTAGTGGA AATGGGGTTC
2901 TTGTCGTCAA CGAAGTAACA GCTCTATTTG GAGAGTGGAG CCAGCCTCAA
2951 AGCCGCCTTG GTTCTATAGT AATCTCTCAG CGTTTGAGTT GACTAATCCG
3001 ATGGTACTTG AAGAAGAGTG AAACGACGAG ATTTCCTTAT AGGAGGAGAA
3051 AGGGATTCCT GTTGCTTTTG GGTGGGTCCA ATCAAGGTAC GAAAAAACTC
3101 TTTTCCCACG GCTCACTAAT ACGATCCTAT CTTACCATTC TTCTATCCTT
3151 CTTCATCACG GGACAGAAGC ATAACGCTGA TCTATGGACG AAGGAAGGTA
3201 TCAGAAAATT CATTATGACC TAGACCTAAA CCATGCTGAT ATTGAACAGA
3251 GGAGGAGTTC TTGTTACTAC TTCTCGCTGG CACCAAAAGA AGTTTCGCCT
3301 GATTCGGACC TTTTGTGGTG GTCTGCCGAG GACAGCACCT AAACAACCCT
3351 TGTGCCATGG AATATTCCCT TTTACGATTG GAGAACCTGG CTTTGATGAC
3401 ATTCCTGAGG AAGGTATTTC CCCTACTAGG CTCGTTGCTT CGCTTATTAA
3451 AGATCAGAGA AGCGAATGCC TCTCGTAATT GATTGAAAAG TGCTCGCCTC
3501 TAGTTCCGGC AGATGCCTGC CATGTACCAT CCGTGGAGGA AGTAGTGGGT
3551 TATTGGAAGT ATAGGATTAT TTATCAAGTG TCACCAAGGA TACACCCCAC
3601 TTACTAAATA GGCCTGTGGA TGCGGCATCT GGCTCTATGC GATGAAAGCA
3651 ATGGGAAATG AGATTCTGGG TTTGATTACT GGTACAAACA AGTTCAAAGC
3701 TAGCGAGAGC ATAAAGAACG GTAGATTCCA TTCAACTGGT AGGTCAAGCC
3751 CTTACCCATA ACACTTATTC CAAAAGGGCT ATCCGAAAGA GGAGATCAAT
                                        XbaI
3801 AATTCCACTA TCAATATGGG TAGGCTGCTA TATCTAGA
```

B.

```
    XbaI
  1 TCTAGAAAGA GAAGATGAAA GGTTTTTTTT TGATTATTAG CAAAGGTACT

 51 TATCTTATTA CTCAGCTAGG CACTTACAAA CCTATTTATT TAAAAGTTCC
                    atp9 promoter
101 ACAGTCAGGA AGACTCGGAA TCATTCTCAC GGCCACCCCA CACGGGAGCG
```


Regulatory Motif
Start site
PstI
atp9
JL
orfC9

1551 AGTTGTTGTT GCTTTTTCTT GGGGAGGGAC AGGTAGGGGG TGGTAGAATA

SmaI

1601 GGATGCGAGA TTCCGATTCG AGGCCTCCCC GGGTTAGTTG TGATTGTGAG

1651 TAATTCTAGT TTAGGCTTCC TGGGGAGGGA TAGAAGAGTG GGCATGTGGG

1701 CTTCTATCAG GGTGGGGGAT TGTCAAAGGT CGATTTCTTC TATTCAATTG

ScaI

1751 GGCCTGCGAT ACTAGCTAAA AGTACTTGAA CCAGGTTAGT CTGCTTGGAA

1801 GGGTAAATGG CATAACTCGA AATGACTAAA GACTAAAGCT AGAACATCTA

1851 TAGTGACAGA TGAGGGCTAA TCGTAACTAG AAGAAGGATA CTACCTCTGA

1901 ACAGCTAAAA ATCAATAAAA TTGTGACCAA GGAGAATTAG ATCAATTCAA

orf1

1951 ACATGAACAC TAATTTTACG AGACGAAGTA GCCGGATTTA TGGTAAGTCC

2001 ACAGCGCTCA CTGAACTCAT GATAAATGTT CAGTTCGAAA ACAGTTTCTT

2051 TCAATCAAAC AACTCTGTAA AGTCATCAGT GGAGTTACAC TTCCGGGGAC

2101 CTCCCTACCC TACTGGGATG GAGAACGACC TCATTTTCAT AGCAATTCTA

2151 TTTATGCTGA AAAAGCATGT GTATTCTCGA ATGACAAATG GATGGACTTC

2201 CATTGTTCCT GTAATGAAAC TCATTACGGG AAAATCATTT CGAACTGAAA

2251 AAGAAGATGT GGTTTATACA GCCCTTTGTC CTTCAATCCA TCTTTACGAT

2301 CTTGGTGAAA AGAAAATCAT ACCTATTCAT GACATTGCAC TGCTAATTTA

2351 TCGTGCAATC AGTATACAGA AGGATAAGTA CCAATCCCAC AGCTTGTATG

2401 GGCTAGCCAT CAATGTATTC AGTGATTCGC TGAACGAAGA TGAAACTGTT

2451 GTTATTCCTA CAGATAATGA GGTTTTTCAT GAAATAGGCA AACTACGGCT

2501 GCTGGAAGAG ATCCCTTCTT CCATAATTGA CAAGAATTCA GATGAATACA

2551 TGGAGGATAA GAAGGACAAA AAATACAATC CAAAATTACC TGATTATATC

2601 ACATCACTCA AAAATCCTAA TCGGAAAATG ATGGATGAAT ATTTAAAGAA

2651 AAAGGAGGCA AATCCACAAT ATAAGGGGAA ACCACCCCTG CCTTCCGTTA

2701 TATTTGCGGC TGATATGGAA ACAATCCAGT TAGAGGCAGA TCCTCGAATT

2751 CAGATTGATG AATCAGTGGG AACTTTTCCC AACCAAAATA AGAGAAAAAG

2801 AGAAAAGCAA AAGTCTCCTC GTCACTATGC TTATGCTGCC TCGTGGATGG

2851 CAGTCAAGCC AGACAAAAAG CTCTCGAAAG CTGACATTCT TCACAGTTTT

2901 TCACATGAGT ACAGGTTCAC TGATCCTGAC TTTTACGATC AGAGCGAAAG

2951 AGTGCTGAGT TCTTTCATCC ATTCACTCGT TAGAGAAGGT AAGAAGCTGA

```
3001 GAATGACACC AATAGTCTAT TTTCATAACT TGTCAAAATT CGATGGTTTT
3051 GTCGTTGCTG AACATTTGTT TACTAAGCAC AATGACGAGT ACGATGTAAT
3101 TTTACAATGC CGTAATAATA CGATATATGA GATATCGGTG TATAAGAAAA
3151 ATCAATTTCG GGGGAAACAA GATTCATTAA AGAGAAAAGA GAAAAGGTTG
3201 CTATTAAGGT TTAGATGTTC ATTACTTATT CTTCCTGATT CGCTCGAGAA
3251 TTTAGCAAAG CAAATATGCC CTGAGTTGGG TGGCAAGGAG GAGATGGATC
3301 ATCAGAGTGT GACTCTTGAT TCTCTTAATG AAGAAGAGAA ATATATTCAA
3351 TATAGGAGAT ATTTATCTCA AGATGTCTTC TTGCTAGGTG TCATTTTGCA
3401 ATCTGTTCAA AAATACTATT GGGACGAATT CCTCATAGAC ATCTCGACAA
3451 AGAAAACAAT CGCATCATTA TCATTAGGGA TATTCCGAAT GGAATACTAT
3501 GATGACATAA AACATCGTAT GTATGTTCCC AACGAGAACG CAGATAAATT
3551 TATAAGGAGC GCTTACTTCG GTGGACACAG CGATGTGTAT AAACCAAAAG
3601 GAGAAAATCT ATATGTCTAT GATGTGAATT CATTATATCC ATCCGTTATG
                    SmaI
3651 GTAGGGAACA AATACCCGGG AGGGAGGCCT GTCTGGCATA GTGATCTAAC
3701 TGAACATAAA CTCGAAGATA TTTTTGGCTT TGTTCAAGCA TTAGTGAAAT
3751 GTGATGGTAT CAACAAACCT TTTTTGCCGA GTAAGCATCC TGTGGAGGAT
3801 ACATTAATCT TCCCAGAAGG AACTTTCTTC GGTACATACT TCACTGAGGA
3851 CCTCAAATAT GCAAAGAGTA TTGGGTACGA AGTTATTATC ACATGTGGAT
3901 ATCTTTTTGA TCGCATGGAC AGCCCATTTG ATGCATTTGT CAATGCTCTG
3951 TATAAAAAGA GAATGAAAGC TAAGGCTGCA GGCAATACAG TAGCAGCATA
4001 TATTTGCAAG TTATTGCTAA ATTCATTATA TGGTAGATTG GCTATCTCAA
4051 CTGAGTGTGA TCAAACTATT GTTCTGAACA AAGAACAAAC AGAGGCTTTC
4101 GTAAGACAAA CACCAACTAA GGATGTGATG GTTCTGACTG CTGATATTAA
4151 TATAATATCC TATATAGAGA GTATGTATGA TACAAAAAGA AAATGGAAGC
4201 CACCAGTGAA TACATGCCCT CAAATGTCAG CAGCAATTAC TGGATACTCA
4251 CGGATCAAAA TGTATCCATA CATATCAAGG GATGACTGCT ATTATACGGA
4301 CACGGATTCA GTGGTTCTTA GCAACCCTCT CCCTGAAGAG GAGGTTTCTG
4351 AGACAGAGCT TGGCAAATTC AAACTTGAGC ATGGAGGAGT AATCACCATA
```


4401 GGATATTTTG TATCACCTAA GTGCTATTTG TTACGGCAAA CCAAGAAAGA
4451 TAATAAGTTT TGTGAGATAG TGAAGTTCAA AGGAGTACCT AAAGGGCTGG
4501 TGCCTCCTTT GTTTTTTGAC AAAATGATCA AAAATATTGA GCATATAGAT
4551 GAGGTCACGT ATACTAACTA TTTTCGTGTT GATTACAAAT TCATGACGGT
4601 TAATATGCGA AAAAGTCATG TTTCACTACG AATGAACTCA AAAAAGAGAA
4651 AAGCAGTATT TGATAGCAAT AATCAATGGA TTGATACCAA ACCAGTTCAA
4701 TTTGATAAAA ATACCATGAG AGCTCAGTTA GCTCGAAGTC GAAACTCAGA
HindIII
4751 AGTCCTGATA CTGTCACTTC TAAGTAGGGA TAAAAGACTT AAGAAAGCTT
4801 TCAATGGTAA AAAAGAATCA CCACCTTCTG ATGATGAGTC AGACTCAGAC
4851 TGAGACTACT GCTGATGAGC CTGCTGTCAT GAATCTCATT TCTATAGTCG
4901 AGAGGTACTA TCTGGATCAT AAACCCATCT GATCCGAATC CCACCTTAAT
4951 CCCATTACAT CAGATTTCAT CTTATACCTT TCCTGTCCTA GTTGTTTTCT
5001 AAAAAACAGT TATCTTATCT GTAGAAGTTG TCTCTCTACA AGCATGGTCA
5051 TGATAAGATG GATATCAATC AATAAGGTCA AACGATATCC GTAAAATCTG
5101 AATGGAAAAT CCTCACCTCA AATGAAACAT TCCAAAATCA AATAAAAACA
orf3
5151 ATCTATTATG AGAAATATTA CTCAATTTCT AAATAGGTTT TACTTGGAGT
5201 TAAATGACGT TTTCTCCGCT AACTGTGATG ATTATGATTC GCTTGAAAGA
5251 GTGATGGTGC CTAAACCAAT TCCAAATAAT GAAACAGAAA ATGATCAAGA
5301 TAGTGCATCC AATGATGGAA GAGGACAAAC AAATATAGAT AATAATTCCA
5351 CTATAATAGA AGGGTGTGAG AAAATAGAAG GGTGTGCAAA CAGTAAGGTG
Sal1
5401 TCGTCGACTC TTATTAAGAT TCCTCGAAGA GAAGAACACA AGAGTTTCAC
5451 CTTCCCCATC CTAAAAATGC TGTATGTTGG AGGAGGAAAA CAATATATAA
5501 CACACCGAAA TATGCCACAT ATGGATTATT ACCATCCTGT TGAGATAAAG
5551 GAAACAGCTT CAAGGTTAGA GTTCTACACT TTGTTCAATA AAAAGCCATT
5601 TTTTGCACGG AATTTCGGAA AGGTGAAATC CATATTTGAT TCACTTGGTT
5651 TGTTAGCTAT AAGGAAGGGA TTTGACTGTA TGACTATAAA GCATATATTA
5701 ATGGAAAAGA CTCATTTAGG TGCGGGTGTC AAAAACCTAT TTCATAATAA
5751 AATCTATTAC AAAACACTGG AAGTAACAAC AATATTCTAT AATCCTGTTT
5801 TGGCTTCATA TGTATACTGT TCGTTTGTAT CAGTAATTCA AACTGGTGTT


5851 GCTTGTACAG TATGGTACAA GCTAATCCCT TGCGGAACTG ACCCTATGAT
5901 ATTCCCCGAT GAGATTAGAT CAATCACAAT TGAGAAATTC ACTCAGGAAT
5951 CACCAGTGGT TGAAGGATAT CAACATATTA CATATCTCCT TACAAACGAT
6001 ATACAGGACA CTGTCTTAGA CTTATTTCGT AATGACGAAA CGTTTACCTC
6051 AATTGTAAAC ATCAACGAAT ACAACCAAGA ATTGTTACAG GAAGAGAAAC
6101 ATCTAAAATG TATGGCAGGA TACTGCACAG CTATTCTCAT TTGTATGTAT
6151 ATCACATTTC AATTAGTGAA TAGTGGAGGT GGATTAGTGG AGGTGGATAA
6201 CCTTAGCCTA TTGTGTTGAG CAAACCAATG ATTACCATTT CTGCTCACGT
6251 AGCTATCTTT AAAGCACAGC TTTTAGTGAA CTTATCTCAA TGACCCACTA
6301 GATTGCACCA GATGGATAAA ACAAAACCAT TTGATAGAAG CCCTCACCCA
6351 CTCTCCTATC CCACCTAACC CCGTGGGGTT TCAGTGGCCT CGTGTTCACT
6401 TTTGAATGAA CACTATACTC GATCTATCGG ATCGTAAATC GAATATGATT
orf2
XbaI
6451 CAGTTAGTGA TGTATACAAT AGTCATACTG GTTTCTGTTC TTTTCTAGAT
6501 GTTTTTTGTA GAAATCGATT TCTAATTCTC CATCTTTGTT TTTGTTTTTC
6551 ATTTCCGCGC TCAACTTATC AACATAGATA GAATAAGCAC TATTGAGGCT
6601 TTTGATCTTT CTATTAAACA AGGTGCTGTC TTTTTTTAAA TTCTTAGATT
6651 TGACACAATA ATCAACACAT GATTTTATCA TAGTTGGGCT TAAGACCACA
6701 TTTAGCAGGA AGTCAGGATA TGTTTGAGTC AAGTCATTCA TTATGACTGA
6751 GCGTGATTGT TCTTTCACTT TACTAGGAGA TGATACAATT CTAAGAACTT
XbaI
6801 TTTCAAGTAA CGTGTCTTGA GTCTCTATAG AATCTAGAGT GTGTGATTTC
6851 TCAATACCAG TAGTATTGTT ATATAACTCG TTGAACATGT TACGACAAAC
6901 TAGAAACTGA GCGTAATCAG CATCACATTT GAATGATTCA CTAGTGATTC
6951 CAAGAATTAA TGCTCTTTCA ATGATATTAT CATATATAAA CTTATTAATG
7001 ATAAGTAAGG GGTGTCCAAA ACTATACAAT GCATTTCTAT AAAATTGTGG
7051 TATTTTATAC GCATGTTCAA TGGTTGTTAT AAAGTTATCA TGAACAGTAT
7101 ATATAGGTAT ATATTCTTTA AGGCTCTGGT TTTCCACTGA GATTATCTCT
7151 CTAATAACAT TGAGAACCAT CACTGCATCC CTCTGATGGA TGAAGTTCGC
7201 AAAGGTGGAG ATCACTGTCT TCCGTGGATC TTTTATATCA GTGTTTTCTT

7251 TCACCGTAAC ACCTGATCTC TTTCTTCTTG TAGAGTCATA AATGACAGAG

7301 ATTATTTGTT CTTTTGTCTT GCAATAGTTT TGTATTGTTT TAAACAAATT

7351 ATTTCCATAG ATTATATTCG AATCTAATTT CGTAGTGATC ATGCAGACCG

7401 ATTGGATCAA TTCCATCAAG TTATTTACAT CTTTATAGTA ATGATCCCAG

7451 AATTGAAAAC AAATTTCAAC TAATTTGTTT ATTGTTGCAT AATCCCTATT

7501 ATGACCAATT TTTTTGATGA AGTCATCACG TGTACTATTT TTGGTTTTAC

7551 CATATATAAT TGGCATATAT ATATTTTTCA TAATTTTTCT GTCAAACACA

7601 GTATTGATTA TATTCAATGC GTGCATGTCT ACATTTTCCT GGATAGGTAT

7651 CGTAGTTGTA TAACATCTTA ATTGATGGAG CAAGCAATCA TATATATCAT

7701 TGATTTCGTT TGAGCCTATT AGGTTTGTAT GCATACCCAT TCTTAGGTCA

7751 AGAAAGAAAA AACTCAAGAT CTGATATGCA CTAGCTGATG CATCTTGTGT

7801 TATAGGGATA TGCTTCATGG TATATGTATC CTTGGTTCTG GCTGCATGAC

7851 AACCCGCTAA AAACTGAAAA GGACTTTTTG CTTGATTGAG TAATGATAAG

7901 ATATCAGTAT CTTCAGTAAA GTTTTCGTTA CTCAAGGGTT TACCTTTATC

7951 CCACTTTTCC AATTTGGTTC GGAACCATTC CGTGGCTGAG GGATAATCCA

8001 TCCATTTATG GAAATGGAAT GAGGTGGCCG CATAAAAACA ATCCAACTGA

8051 TCTTGGGTTA ACTCCTCCTC TTCTTTTTTG GTGAAATCAT TCATTTTAGC

8101 TTGAATACGA TTATTTGGTG TTGCAAAGAG GAATAGCGAT CTCATGAAAT

8151 CTCTCTCGTG TAGATGAAAC ACTCCACGTC GATAATTTCT ACCACGGAAA

8201 TCAAGAAATA ATGGAAGATA GAATTTATAA CCAACCAATG ATTCAGCAAG

8251 TTCCATTGTT GCTTGTTCAA ACATGGCTTT CTGAACGTTT TTACGGAACA

8301 GATTCATCAT ATCATAAATT TTACAGTGAT CTCTCATTTT TTTGAATCTA

8351 GCACTACTAT AGAGATCTCT TAGCTTATTT ACACCATCAT TTATGTTGAT

8401 ATACGCCAGA CTCACAGGCA TGACAAGTCC AATCCTCTGT AAATCACTCC

8451 ATTTACACTG TAAATACTTC AGGAAGTTGA CATTGATGGT AAATGGCTGC

8501 TCCTGTAGCC TGTTCAAGGC TTGGGCAATA TCTTTGCCAT CATCATTATT

8551 GCAGTTAGTT TGAATATCAA ATTGATCAGA ATCATGACTA CTTATGAGAG

8601 AAACGTCACT CTTCATCCTA CCTTTGATTG CTCCTGTGAA GTACCCTCCT

8651 ATCATATTTG ATATACTTCG AATCTTGACA GTGTCCCATT CTGGCTGGGC

```
 8701 TTTGCCTGTG TTTTTAATCT CATTGAGATG CTCCAATAAT TCCTTTATGA
 8751 GGACAGGGTA CACATTTCGT GTGTAGTCTG ATTCTTGACG GTAATCCATC
 8801 TGCCATCGGC GTGGTTTATC TACCATGGGT AGGGGTACAC TGATTGGAAT
 8851 ATCATTTTCC TCGAAATTAC ATGTGACATA GGACACAGGA CTCCTGTATT
 8901 TCACTTTACC TGTTTTTTTA TCCTTGCTAA ATATAGCTGG TCCAGGGGGA
 8951 ACCTGCTTCA GCAAGTCCTT CTTCGCTAAG TAATCAAACA TCTGTGTCCC
 9001 TATAGCATAA GTTTTATTTG CTCTCTTTTT TGCTTTCTTT TCTTCTTGAT
 9051 TTAGAATAGC ATCTTGTTTT AGAGTAGAAT CTTGTTTTAG AGTAGAATCT
 9101 TGATTTAGAA TAGCATCTTG TTTTAGAGTA GAATCTTGTT TTAGAGTATC
 9151 ATCTTGATTT TGAGTAGAAT CTTGATTTAG AGTATCATCT TCGATTTTCT
 9201 TTTTGCTTTT ATTAAGATAA TAGGAAATCG CTTTTGATCC TATTTTCTTT
                          BamHI
 9251 GCCTTAATAT AATCTGGATC CGACATCATG GCAATATGAG CCCTGACTGT
 9301 CAAATCAAAT TCTTTGAACA AAGTCTGACA ACTTACTCGT TTGTACTCCA
 9351 AATTAAAAGG ATTAAAGTGT CTACTTAGCA CCCAAACACA AATGCATTCC
 9401 AAGGTAAAAG CTCCCAATTG CTTTACAATT TGTATATCTT CCTCGTTTTT
 9451 TAATTTTTTC AAAAGAACAT CCTTAAGAGT GATTTCATTC TTTATTTTTT
 9501 TCTTATAGCC TTCTTTGTTG TCTTTGTCTA AGATAGTCTT TTTAACTCCT
 9551 CGGATTTGAT CAATCAAGTG TTTGAACTTA ACATGAACTG ACTTTTCACT
 9601 AAAGTCAATA GTGAGGTCTT CTACATCATA TTGAAGCTGA TGGAAATCAA
 9651 TCCCGCAATT CATACTGAAA TCATCAATTT GATGATTATA GAATGACTTA
 9701 TAATCAAAAT TCGATAATGC TTTTTTAGTT TCTTTCCAAT AAGTAAAAGA
 9751 AGCATATTTG CAAAGATTTT TGTTTTGTTT ATTATCACGG ATATTTACAC
 9801 TAGCTAAAAG TGTATCCCAG AAGGCATATA TTAAATTATA ATCATCCACA
 9851 GTGAATTTTG TTCCATCCAA TACTTCTTTA GGTATCCAAT ACTGTTTATT
 9901 TTGTGCAGTA TTATATAGAC TATCATACTG GATAAAATAT CCATTTTTTC
 9951 GAGCATATTC ATCTATTTCT TCAAGTTTAT CAACAAGTTG ATTAATTTCA
10001 TCAAAATCAA AGTGTAGATA TAAATCAAGT CTTATTTCTT TCTCATGTGT
10051 TGTTTTCACA TATTCTCTTA CTTTATTATT TGTTGGCTTT TCTGCATGTA
```

```
10101 AAGCAAACCT TATCTGAGCA GCCTCTTTTG TGTCTCTGAT TAAATCCAAC
10151 CTTTCATTTT CAAGTTGCTT GAGTTTATTA TAAACTTCAA TGTTTTCACA
10201 ATTTTCTTCC ATTGGATGGA ATCCAGCTAG GGGTTGTAGC AGTAAATTAT
10251 CATTATGAAC CATTTGAGAA AATGTCGCAT CTCCTTGGTC TGTCTTTTTA
10301 ATATCCATCT ATCACGATTT TCCAATTCTT GGTTTAGTTC ATCTGTGTTG
10351 AATCCCTAAA TCAAACATGG GGTTCTACGT GAGCTCCGTT ATGGAGACTC
10401 TCCATGAGTA TTCACATCAT AAAGTCAACA TTAGAGTCGA CTCTAGTAAC
                  PvuII
10451 TCAGGACGAA CCAGCTGGAC AAAAAACAGG ACCGATAGGA TGTATCATTA
                                                      JR
10501 TCTTCTTTCT GGCTTAAAAG TCGGGTAAAT GGTTGTGGTA CTTCAATCGT
10551 AAGTTGATTC AAGTCAGCGA GAAAGAAGGG AATTGCTTGG TTGATAGGGC
10601 CCCGCCCTTC GAGTTTGTCT CCATCCCGTA CTGATTAGAA AGATTTACGC
10651 GGTTAGATCG AAATTACTTA TTATCGACAC GGATTGGTAA CTCCTTCTAC
10701 TGATAGAAAA GAAAAGAACA AGCATATTGC TTGCAAAGAC TCCCACCCTA
                                                      SmaI
10751 TTCTTATTAG TAGGCAAGGG TCACTCACAA CGGATATCAT TATGGCACCC
10801 GGG
```

C.

```
      SmaI
    1 CCCGGGCAGT ACGAGATCAG GGAGCCTTCC AAAGGGCTCT ATTACCTAGA
   51 TTGGATAGTC AGAAGGGGCG CCCCACCCAA TGTGAATTGA GAAGTATTAA
  101 AGATCGTATA GAGACGATAG ATCGGTTCGT ATCGTATTTC TAAAGATGGC
  151 GCTGAGCACA ATAGCAAAAG GGTTAAAGGA TTTTGCGTGG GCTTAAAATA
  201 TCCCATTACT TCCTCTAGCT CTTTCAAAAG GGATTAGAAC GTAGACTGAC
  251 TACTACAAGA AAGGGTACGA GGTATAGATA GAAATAGGGG TACTTACGAA
  301 CCAAGGTAAC TGATGCAAAA ACGGACCACC CATACTCCTT GATAACTCCT
  351 TTTTTCCTGC CCAATCTAGC AACCCACAGA GCGGTAGACA ACCGAGGCAA
  401 CAAAGACTAT GATTATCCTT TAGAATTGAA CCCACTTAGG CTGGTTCATT
  451 CAACCATTTT TGATTCCATC CTTCTTCTAC GAGAAGTGAA AAGAGGCTGT
  501 GTCATGACGA CAGCAATGCC GATAGTCATC CGTCTTATCC GTCAACTTAA
  551 GCAGACTAGT GGTAGTCGAA CGAATAGCAA TCCATGCTGC CAACTAGTAA
```

```
 601 ACCAAGAGGT TTTGTTTGGT CGAAAGTCTC CTCCGCTTCG TCTTTGTACC
 651 ACCGCATTCT TTGCTTTGCC CGTATCATAT GATCGCGAAG TGTGCTTCGC
 701 AACAAGATTT GTATTGTACC AAGGACTGGA AAAAAAGATC AAAAAGGTAT
 751 CAACATCGGC TTTTTTCTCG AAACGAAATA GACTAATTGG ACTGAAAATC
 801 TGCACTTCCA AATTATAGAA TCTATTTTTC ATTAATTTGT TGAGTAGTCA
 851 TTTTTTTTTT ATATCGTCGT GTATCATCTA ATTTACTTAT TCCCGAATCT
 901 AAGGACTGGG GAGTTCTAAT CTCGGGAAAT GTCCCAGTTC AACGAGAGGA
 951 ATCCAACTCA GCCGGTCACA CAGACTTCGT TCGAATTACG GATTTTATGA
1001 TGAAAAAAGA AATGAAATTC AGTAGAAGAA GAAGGCTTCT CGTCGCTCCG
1051 CTTATTGCTG CCACGCAAGA TTGATGGAAA GCTCGGTGAG TGAGTCGTTC
1101 TTACTTGGCA CAAGCGTTGT CTCAGCTAGG CATCCTTCTA TGATGTGACA
1151 GATTGGTTAG GAGGTTAGTC GTAGAGCCAG CCCCTGAGTC CACTTCCCAT
1201 TGTCATAGAA AACCAGATGA CGTAACGTAA TAGGAAAGTC TGCCGCCGAG
1251 TCAGAGTACT CAAGGAGTAG CAAGGAGAAG TAAGGAAGTA GGCGCCGAAG
1301 TAAAAGCTAG ACCAGCATCT TCTCCTATTG AAATTCTTGG AGAGAAAGGA
              HindIII
1351 AAACGGAAAA GCTT
```

D.

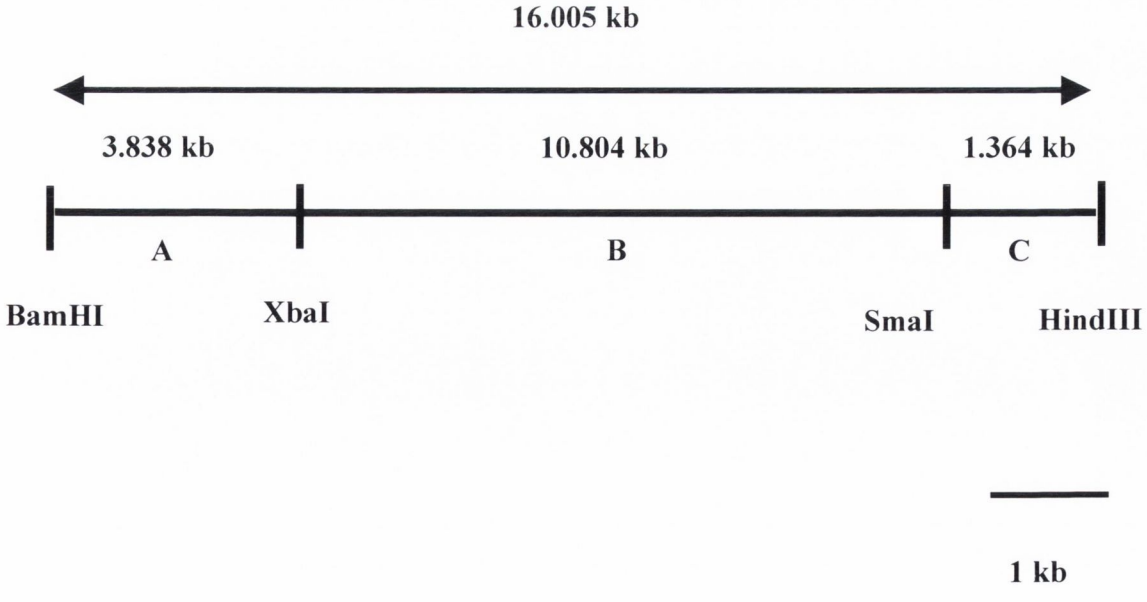


Figure 4.10 Nucleotide sequence of a 16.005 kb region in mitochondrial genome for the sterile line of *L. perenne* CMS9B290.

A. The 3.838 kb nucleotide sequence is shown, restriction sites *Bam*HI, *Xba*I. and *Hind*III is underlined. The green coloured sequences is identical with the fertile LPS21sequence. **B.** The position of the open reading frames orf1, orf2, orf3 along with the chimeric *orfC9* gene and the known *atp9* gene in the 10.803 kb is highlighted by boxes. The atp9 promoter is underlined and start site is shown by the arrow symbol ▼. The junction sites **JL** is indicated, starting at position 980 bp and **JR** is indicated starting at position 10.539 kb. The restriction sites *Xba*I, *Bam*HI, *Sma*I, *Pvu*II, *Bam*HI, *Sal*I, *Hind*III are underlined. **C** 1.364 kb is single strand sequenced from *Sma*I to *Hind*III 10.804 kb. **D** is a sequencing map showing the position of the region A, B and C in CMS9B290 for *L. perenne*.

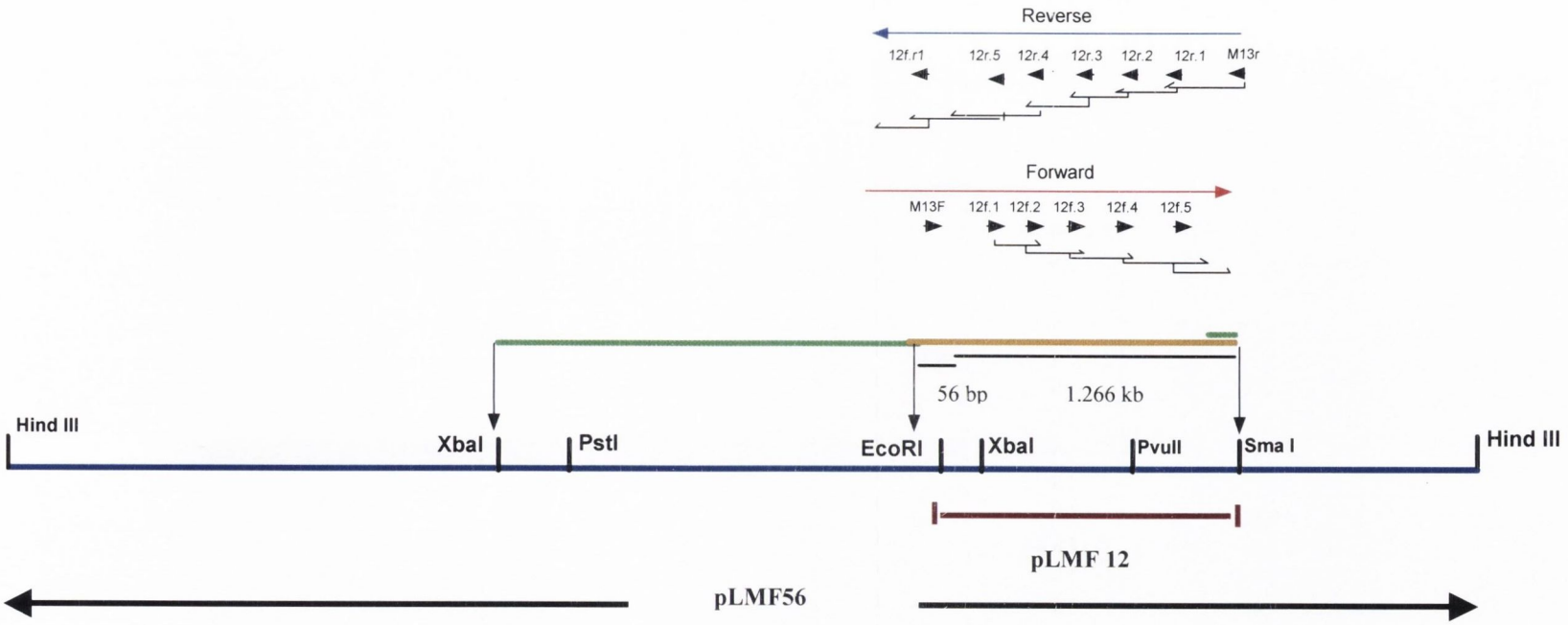
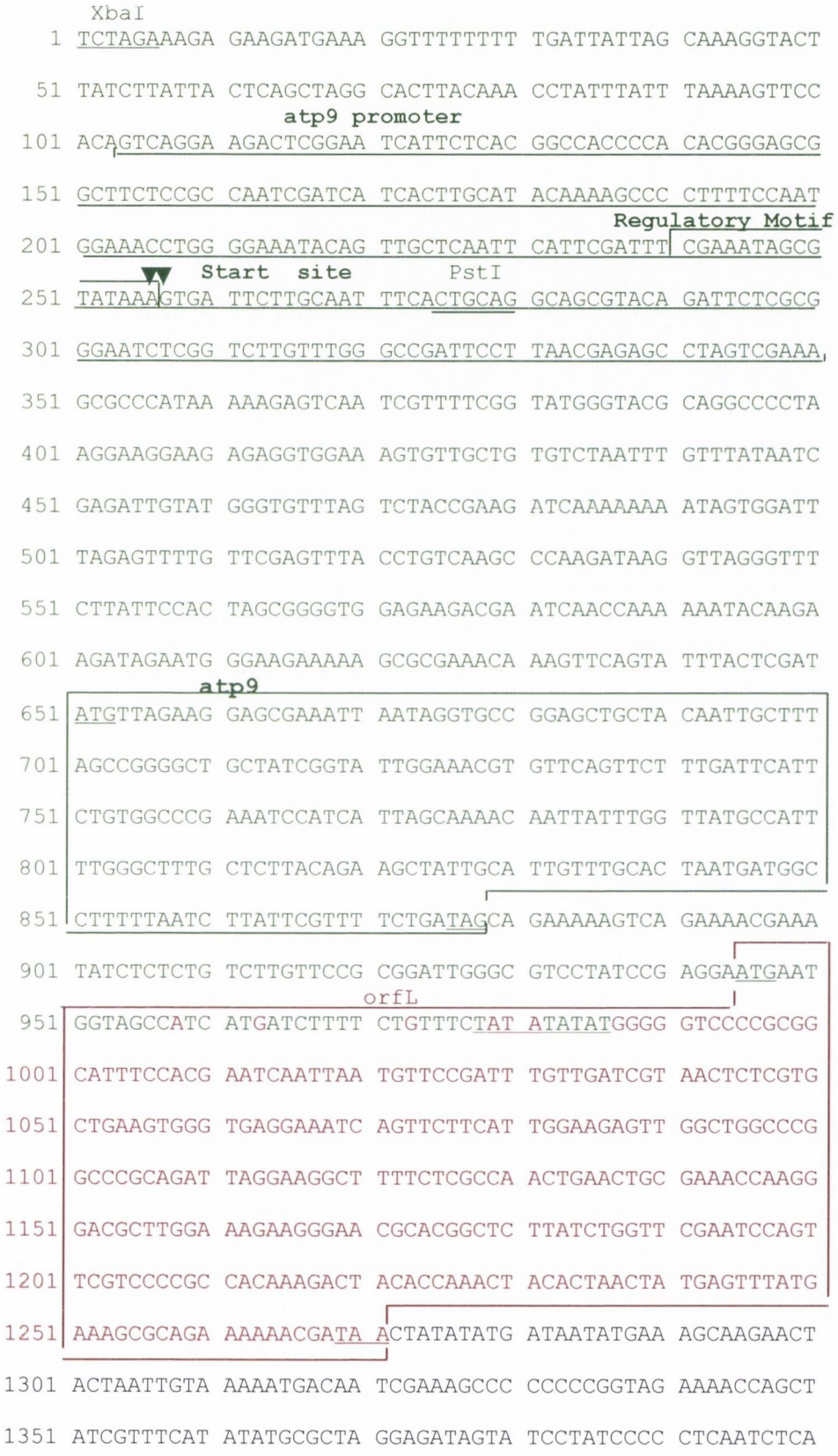


Figure 4.11 Restriction map of the fertile clone pLMF56 and sequencing map of the clone pLMF12

The primers designated for use in sequencing are represented by the arrow symbols ➔ . The region sequenced by previous workers (Kiang and Kavanagh, 1996) is represented by the line ▬ additional sequence is indicated by the line ▬ .


XbaI
TCTAGAAAGA GAAGATGAAA GGTTTTTTTT TGATTATTAG CAAAGGTACT
TATCTTATTA CTCAGCTAGG CACTTACAAA CCTATTTATT TAAAAGTTCC
atp9 promoter
ACAGTCAGGA AGACTCGGAA TCATTCTCAC GGCCACCCCA CACGGGAGCG
GCTTCTCCGC CAATCGATCA TCACTTGCAT ACAAAAGCCC CTTTTCCAAT
Regulatory Motif
GGAAACCTGG GGAAATACAG TTGCTCAATT CATTCGATTT CGAAATAGCG
Start site
PstI
TATAAAGTGA TTCTTGCAAT TTCACTGCAG GCAGCGTACA GATTCTCGCG
GGAATCTCGG TCTTGTTTGG GCCGATTCCT TAACGAGAGC CTAGTCGAAA
GCGCCCATAA AAAGAGTCAA TCGTTTTCGG TATGGGTACG CAGGCCCCTA
AGGAAGGAAG AGAGGTGGAA AGTGTTGCTG TGTCTAATTT GTTTATAATC
GAGATTGTAT GGGTGTTTAG TCTACCGAAG ATCAAAAAAA ATAGTGGATT
TAGAGTTTTG TTCGAGTTTA CCTGTCAAGC CCAAGATAAG GTTAGGGTTT
CTTATTCCAC TAGCGGGGTG GAGAAGACGA ATCAACCAAA AAATACAAGA
AGATAGAATG GGAAGAAAAA GCGCGAAACA AAGTTCAGTA TTTACTCGAT
atp9
ATGTTAGAAG GAGCGAAATT AATAGGTGCC GGAGCTGCTA CAATTGCTTT
AGCCGGGGCT GCTATCGGTA TTGGAAACGT GTTCAGTTCT TTGATTCATT
CTGTGGCCCG AAATCCATCA TTAGCAAAAC AATTATTTGG TTATGCCATT
TTGGGCTTTG CTCTTACAGA AGCTATTGCA TTGTTTGCAC TAATGATGGC
CTTTTTAATC TTATTCGTTT TCTGATAGCA GAAAAAGTCA GAAAACGAAA
TATCTCTCTG TCTTGTTCCG CGGATTGGGC GTCCTATCCG AGGAATGAAT
orfL
GGTAGCCATC ATGATCTTTT CTGTTTCTAT ATATATGGGG GTCCCCGCGG
CATTTCCACG AATCAATTAA TGTTCCGATT TGTTGATCGT AACTCTCGTG
CTGAAGTGGG TGAGGAAATC AGTTCTTCAT TGGAAGAGTT GGCTGGCCCG
GCCCGCAGAT TAGGAAGGCT TTTCTCGCCA ACTGAACTGC GAAACCAAGG
GACGCTTGGA AAGAAGGGAA CGCACGGCTC TTATCTGGTT CGAATCCAGT
TCGTCCCCGC CACAAAGACT ACACCAAACT ACACTAACTA TGAGTTTATG
AAAGCGCAGA AAAAACGATA ACTATATATG ATAATATGAA AGCAAGAACT
ACTAATTGTA AAAATGACAA TCGAAAGCCC CCCCCGGTAG AAAACCAGCT
ATCGTTTCAT ATATGCGCTA GGAGATAGTA TCCTATCCCC CTCAATCTCA

1401 GGAGTAGCTT TCTCTATGCC CGACTTACGA GCTCTTTGCT TATTATTTCT

1451 CGGGTGGGAT GAAGGTCGAA ATGTCTAGTT TGAATTGCTT ATGCTTATAT

1501 AGTTTGAATC GCCTAGCTTC ATACCAAGGT TGTTCTACAA AAGAACCAAT

1551 AAGAAAGTAA GCCCGTTCAA TTGAAATAGC TTCCTCGCTT CGCAGTTCGA

1601 TACCCACTGA AACTATGGAT ACGATAGACT CGGGAGGGGG AAAGGGTTAC

1651 TTTATTTGGG CTGCTAAGTC TTTCTTGTTG ATTCCGGCAA TAAAGAGGAC

1701 CTTCTTAATC CTTCGTGCTC CACACCCGCC TCACTGGACC TATCCATTCG

1751 AGCTCTCTCC TCTAATCCCA ACTCGACACG GCTGGACGAC GACGCCTGTT

EcoRI

1801 CACTTATAAT GAAATCACAC GTTAAGCCCG ATGAATTCAT CTACAATTTA

1851 TTGATTTGTC GGTTTGGCAA GGCAGAACCG GGGCCCTTCC ATTAAGGTTA

1901 TATTAGAGTT TCGATAGGAG AGATGCAGAC CAGATAGGAA TGCAAGAATA

1951 CTCTGTAAAT TGTTTATTAA TTGGTTTAGT ATGTGGAAGA GAAGTATCGA

XbaI

2001 TCTAGAGGAT TGCCCCCCCA GACTGACCTA GTTGTCCTTG ACTTGCTTCG

2051 GAAAGAAAAG ACCTAGTACT TGTTTACCTT GACCATGGCT ACTTTCCCGC

2101 CGCCTTTCCT TCTATACTTT TCCTATGTGC TATCCCAGGC ATCCCTTGTC

2151 AAGCTACCTA TCCCTACGAA AAAGCAAGCC TACTTACTAC GTGTTTTCCT

2201 CTATTGGCAA ATACGCATAT ACTGACCTAC TCACGTGCAT ATCTTGCATA

2251 AAGCCTTCCC CTAAAAAAAC AAACTAGCCC GTATCTTCCT AGTCCTTGTT

2301 GTCGTTGAAT CTGGCGCTTT AGGGGCAGAA TGTGTGATTG ATTCCAGCGA

2351 GAATACATAA TCTCTTTTTC CGCTGCATAG CCCGAACAAC ACCAATATCG

2401 GGACACCCAT TATAAGAAAT CAGCAACTCG AATAATCGTA CCAAACGGGA

2451 TGTCATTCAC TAGATCGAGC AATAGTAACA AAGCGAGTAA GAGGTTTTTT

2501 CCTAACAATG CGTGGGGGTA TTTACTTCTC CATGCGTCAT GAAAATGGGA

2551 TAAAGAATCC AATTGGAATC GAGGGCTCGA GTCTAGTGGA GTAGTGAGGG

Pvu II

2601 TAGCAAAGCA GCTGTTACCC TTGTATGGTT AATAGGAAGG AATGCTCTTT

2651 TCAGCCAAAG ACTACTAATT TCTCATTTAT TAAATTGGGT GAAGGCCTGA

2701 CGTGTAATCT ATCGTTAAGG GTGCCTATCA GAAATTATCA CCTTTAGAAA

JR

2751 GTAGCCTTAG AATCGAATCA TCTTGTCAAT CCTCACTTCA ATCGTAAGTT

2801 GATTCAAGTC AGCGAGAAAG AAGGGAATTG CTTGGTTGAT AGGGCCCCGC

```
2851 CCTTCGAGTT TGTCTCCATC CCGTACTGAT TAGAAAGATT TACGCGGTTA

2901 GATCGAAATT ACTTATTATC GACACGGATT GGTAACTCCT TCTACTGATA

2951 GAAAAGAAAA GAACAAGCAT ATTGCTTGCA AAGACTCCCA CCCTATTCTT
                                                  Sma I
3001 ATTAGTAGGC AAGGGTCACT CACAACGGAT ATCATTATGG CACCCGGG
```

Figure 4.12 Nucleotide sequence of a 3.048 kb region in mitochondrial genome for the fertile line of *L. perenne* LPS21

The position of the open reading frame *orf-L* and *atp9* along with the atp9 promoter is highlighted. The position of the junction sites JL and JR are indicated. The green colour represents nucleotide sequence homologous with the sterile line CMS9B290. The restriction enzyme sites *Xba*I, *Eco*RI, *Sma*I, *Pvu*II, *Pst*I are highlighted, the atp9 promoter is underlined along with the start site indicated by the arrow symbol▼. The 225 bp *atp9* gene and the 326 bp *orf-L* fertile chimeric gene are enclosed in boxes.

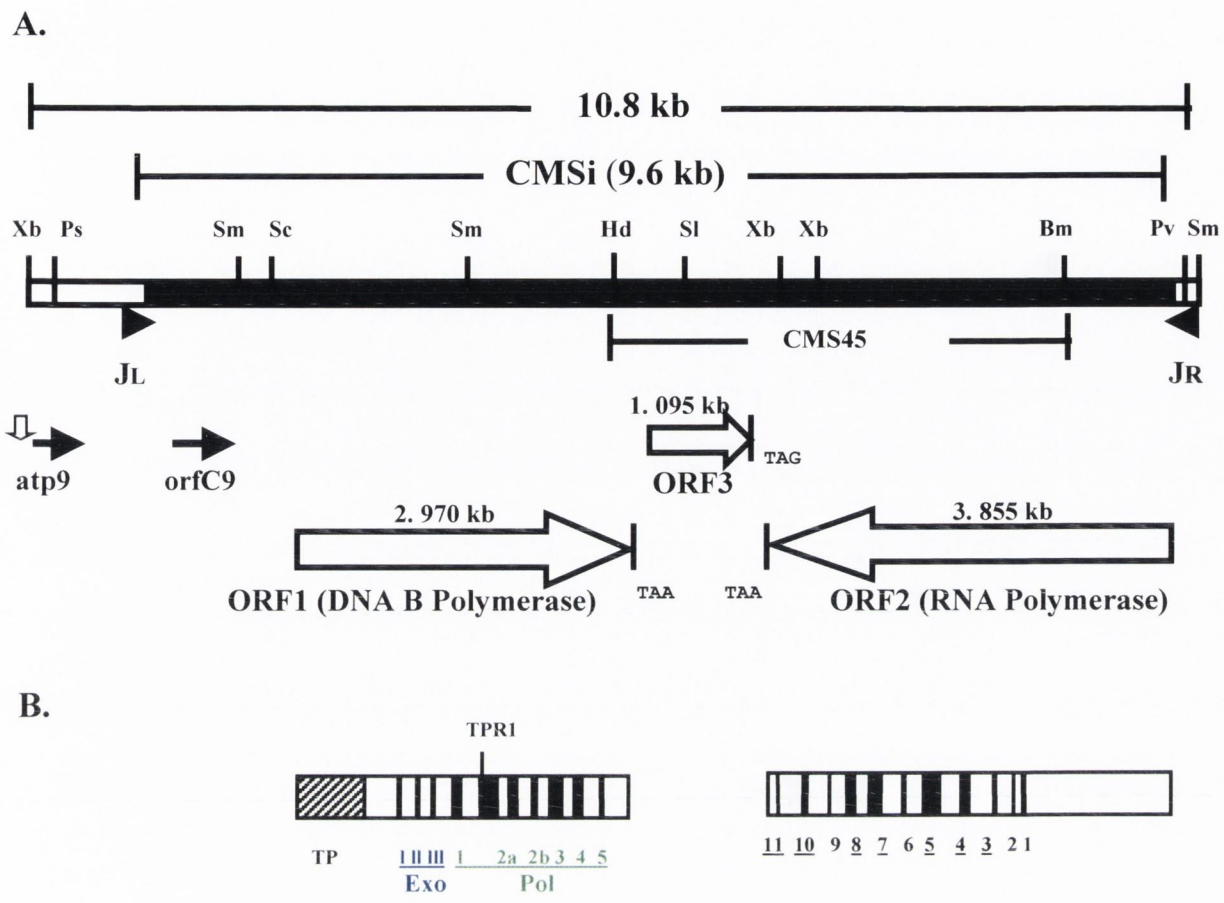


Figure 4.13 Illustrates a 10.8 kb region in the mitochondrial genome of the sterile line of *L. perenne* (CMS9B290)

A. In the line CMS9B290, the open box is homologous sequence to a fertile line of *L. perenne*, and consists of an *atp9* gene with promoter (*open arrow* ⇩). The dark coloured box refers to an integrated 9.6 kb CMSi associated region, comprising the arrangement of a putative linear type plasmid and a chimeric *orfC9* gene. Restriction sites and junction sites are shown by; *Bm- Bam*HI, *Hd- Hind*III, *Pv- Pvu*II, *Ps- PstI, Sl- Sal*I, *Sc- Sac*I, *Xb - Xba*I, and JL ►, JR ►, respectively. Predicted plasmid ORF's are represented by the *unfilled horizontal* arrows and *solid horizontal* arrows represent the ORF'S for *atp9* and *orfC9*. Indicated is the CMS45 fragment previously used by workers as a diagnostic probe in Southern hybridisation to screen for lines of *L. perenne* containing the CMS cytoplasm (Kiang and Kavanagh, 1996). **B.** Conserved domains for ORF1 and ORF2 are highlighted by filled boxes and are representative of DNA and RNA polymerases respectively. The Putative TP (terminal protein) domain for ORF1 is highlighted by a *hatched box.*

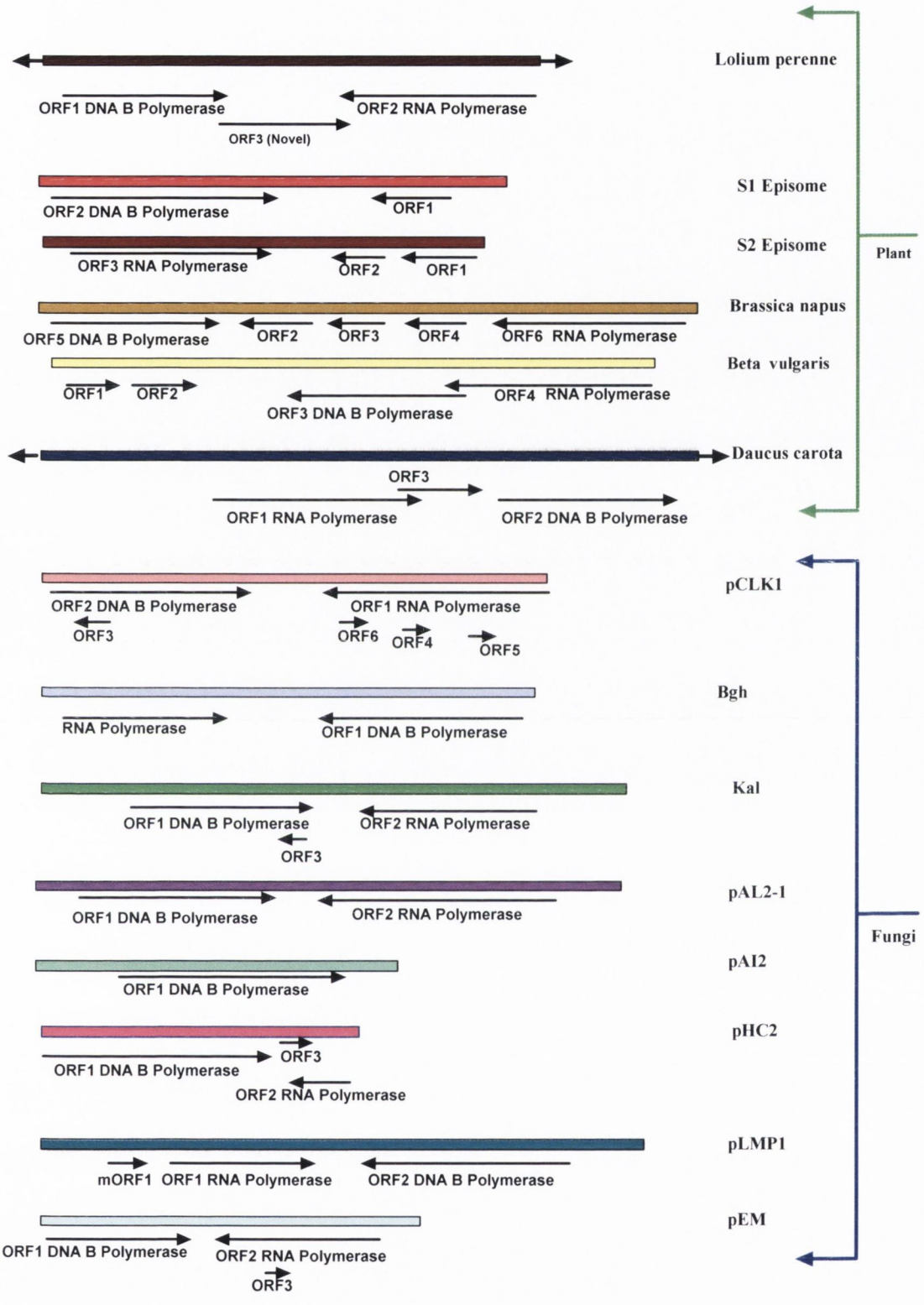
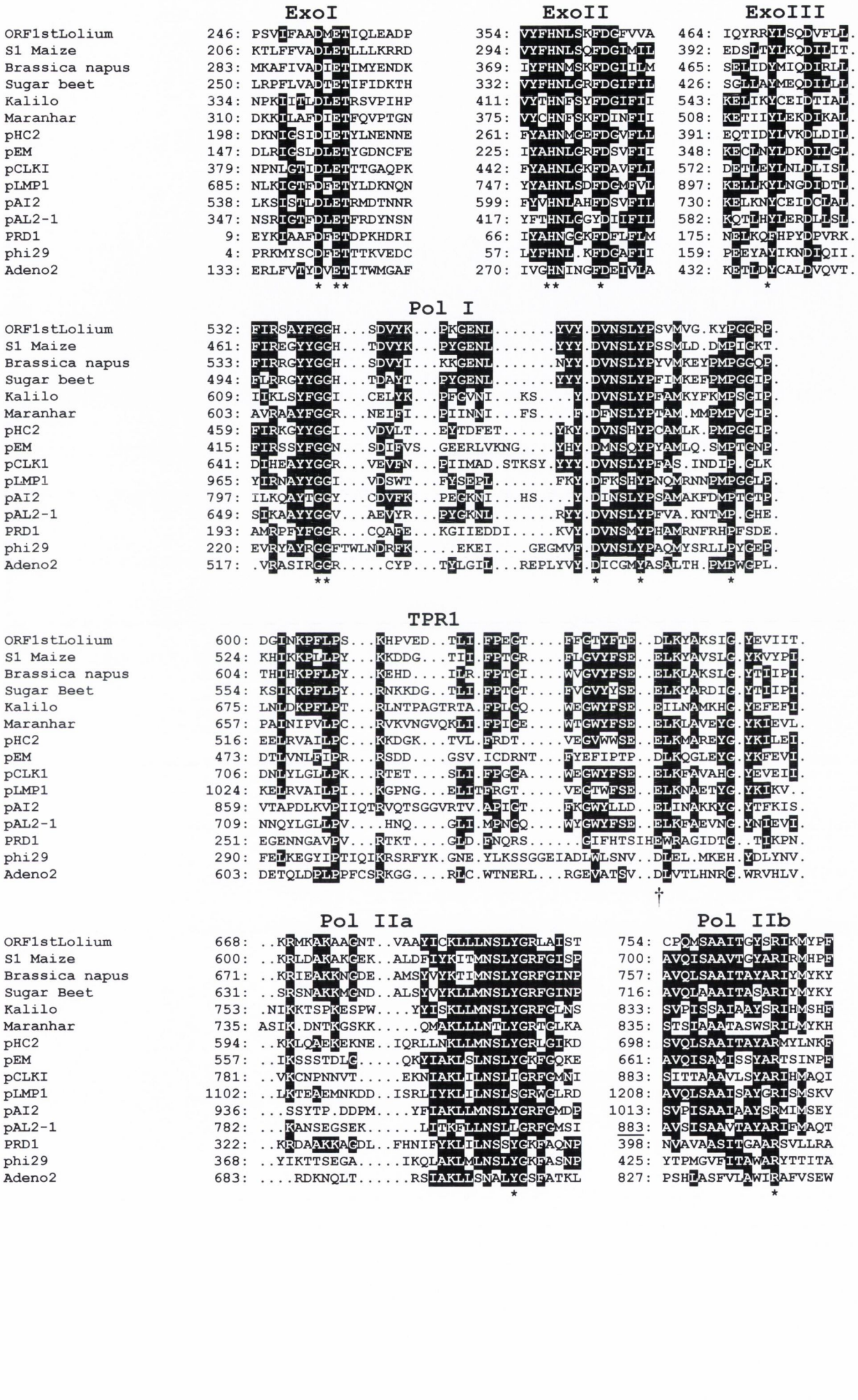


Figure 4.14 Arrangement of DNA Polymerase and RNA Polymerase in linear mitochondrial plasmids compared to ORFs in the 9.6 kb-CMSi element in *L. perenne*.

The linear mitochondrial plasmids in plant and fungi are represented in each case by coloured box, and the location of open reading frames (ORFs) are indicated by a black arrow. All of the linear plasmids shown are free living except for putative linear plasmid like sequence in *L. perenne* and *D. carota* being integrated into the mitochondrial genome.


ExoI
ExoII
ExoIII
ORF1stLolium 246: PSVIFAADMETIQLEADP 354: VYFHNLSKFDGFVVA 464: IQYRRYLSQDVFLL.
S1 Maize 206: KTLFFVADLETLLLKRRD 294: VYFHNLSQFDGIMIL 392: EDSITYLKQDILIT.
Brassica napus 283: MKAFIVADIETIMYENDK 369: IYFHNMSKFDGIILM 465: SELIDYMIQDIRLL.
Sugar beet 250: LRPFLVADTETIFIDKTH 332: VYFHNLGRFDGIFIL 426: SGLLAYMEQDILLL.
Kalilo 334: NPKIITLDLETRSVPIHP 411: VYTHNFSYFDGIFII 543: KELIKYCEIDTIAL.
Maranhar 310: DKKILAFDIETFQVPTGN 375: VYCHNFSKFDINFII 508: KETIIYLEKDIKAL.
pHC2 198: DKNIGSIDIETYLNENNE 261: FYAHNMGEFDGVFLL 391: EQTIDYLVKDLDIL.
pEM 147: DLRIGSLDLETYGDNCFE 225: IYAHNLGRFDSVFII 348: KECINYLDKDIILGL.
pCLKI 379: NPNLGTIDLETTTGAQPK 442: FYAHNLGKFDAVFLL 572: DETLEYLNLDLISL.
pLMP1 685: NLKIGTFDFETYLDKNQN 747: YYAHNLSDFDGMFVL 897: KELLKYLNGDIDTL.
pAI2 538: LKSISTLDLETRMDTNNR 599: FYVHNLAHFDSVFIL 730: KELKNYCEIDCIAL.
pAL2-1 347: NSRIGTFDLETFRDYNSN 417: YFTHNLGGYDIIFIL 582: KQTIHYLERDLLSL.
PRD1 9: EYKIAAFDFETDPKHDRI 66: IYAHNGGKFDFLFLM 175: NEIKQFHPYDPVRK.
phi29 4: PRKMYSCDFETTTKVEDC 57: LYFHNL.KFDGAFII 159: PEEYAYIKNDIQII.
Adeno2 133: ERLFVTYDVETITWMGAF 270: IVGHNINGFDEIVLA 432: KETIDYCALDVQVT.
Pol I
ORF1stLolium 532: FIRSAYFGGH...SDVYK...PKGENL.....YVY.DVNSLYPSVMVG.KYPGGRP.
S1 Maize 461: FIREGYYGGH...TDVYK...PYGENL.....YYY.DVNSLYPSSMLD.DMPIGKT.
Brassica napus 533: FIRRGYYGGH...SDVYI...KKGENL.....NYY.DVNSLYPYVMKEYPMPGGQP.
Sugar beet 494: FLRRGYYGGH...TDAYT...PYGENL.....YYY.DVNSLYPFIMKEFPMPGGIP.
Kalilo 609: IIKLSYFGGI...CELYK...PFGVNI...KS....Y.DVNSLYPFAMKYFKMPSGIP.
Maranhar 603: AVRAAYFGGR...NEIFI...PIINNI...FS....F.DFNSLYPTAM.MMPMPVGIP.
pHC2 459: FIRKGYYGGI...VDVLT...EYTDFET.....YKY.DVNSHYPCAMLK.PMPGGIP.
pEM 415: FIRSSYFGGN...SDIFVS..GEERLVKNG....YHY.DMNSQYPYAMLQ.SMPTGNP.
pCLK1 641: DIHEAYYGGR...VEVFN...PIIMAD.STKSY.YYY.DVNSLYPFAS.INDIP.GLK
pLMP1 965: YIRNAYYGGI...VDSWT...FYSEPL.....FKY.DFKSHYPNQMRNNPMPGGLP.
pAI2 797: ILKQAYTGGY...CDVFK...PEGKNI...HS....Y.DINSLYPSAMAKFDMPTGTP.
pAL2-1 649: SIKAAYYGGV...AEVYR...PYGKNL.....RYY.DVNSLYPFVA.KNTMP.GHE.
PRD1 193: AMRPFYFGGR...CQAFE...KGIIEDDI.....KVY.DVNSMYPHAMRNFRHPFSDE.
phi29 220: EVRYAYRGGFTWLNDRFK.....EKEI....GEGMVE.DVNSLYPAQMYSRLLPYGEP.
Adeno2 517: .VRASIRGGR.....CYP...TYLGIL...REPLYVY.DICGMYASALTH.PMPWGPL.
TPR1
ORF1stLolium 600: DGINKPFLPS...KHPVED..TLI.FPEGT....PFGTYFTE..DLKYAKSIG.YEVIIT.
S1 Maize 524: KHIKKPLLPY...KKDDG...TII.FPTGR....PLGVYFSE..ELKYAVSLG.YKVYPI.
Brassica napus 604: THIHKPFLPY...KEHD....ILR.FPTGI....WVGVYFSE..ELKLAKSLG.YTIIPI.
Sugar Beet 554: KSIKKPFLPY...RNKKDG..TLI.FPTGT....PVGVYYSE..ELKYARDIG.YTIIPI.
Kalilo 675: LNLDKPFLPT...RLNTPAGTRTA.FPLGQ....WPGWYFSE..EILNAMKHG.YEFEFI.
Maranhar 657: PAINIPVLPC...RVKVNGVQKLI.FPIGE....WTGWYFSE..ELKLAVEYG.YKIEVL.
pHC2 516: EELRVAILPC...KKDGK...TVL.FRDT.....VEGVWWSE..ELKMAREYG.YKILEI.
pEM 473: DTLVNLFIPR...RSDD....GSV.ICDRNT...FYEFIPTP..DLKQGLEYG.YKFEVI.
pCLK1 706: DNLYLGLLPK...RTET....SLI.FPGGA....WPGWYFSE..ELKPAVAHG.YEVEII.
pLMP1 1024: KELRVAILPI...KGPNG...ELITFRGT.....VEGTWFSE..ELKNAETYG.YKIKV..
pAI2 859: VTAPDLKVPIIQTRVQTSGGVRTV.APIGT....FKGWYLLD..ELINAKKYG.YTFKIS.
pAL2-1 709: NNQYLGLLPV....HNQ....GLI.MPNGQ....WYGWYFSE..ELKPAEVNG.YNIEVI.
PRD1 251: EGENNGAVPV...RTKT....GLD.FNQRS.....CIFHTSIHEWRAGIDTG..TIKPN.
phi29 290: FELKEGYIPTIQIKRSRFYK.GNE.YLKSSGGEIADLWLSNV..DLEL.MKEH.YDLYNV.
Adeno2 603: DETQLDPLPPFCSRKGG....RLC.WTNERL...RGEVATSV..DLVTLHNRG.WRVHLV.
Pol IIa
Pol IIb
ORF1stLolium 668: ..KRMKAKAACNT..VAAYICKLLLNSLYGRLAIST 754: CPQMSAAITGYSRIKMYPF
S1 Maize 600: ..KRLDAKAKCEK..ALDFIYKITMNSLYGRFGISP 700: AVQISAAVTGYARIRMHPF
Brassica napus 671: ..KRIEAKKNCDE..AMSYVYKTIMNSLYGRFGINP 757: AVQLSAAITAYARIYMYKY
Sugar Beet 631: ..SRSNAKKMCND..ALSYVYKLLMNSLYGRFGINP 716: AVQLAAAITASARIYMYKY
Kalilo 753: .NIKKTSPKESPW....YYISKLLMNSLYGRFGLNS 833: SVPISSAIAAYSRIHMSHF
Maranhar 735: ASIK.DNTKGSKK.....QMAKLLLNTLYGRTGLKA 835: STSIAAATASWSRILMYKH
pHC2 594: ..KKLQAEKEKNE..IQRLLNKLLMNSLYGRLGIKD 698: SVQLSAAITAYARMYLNKF
pEM 557: ..IKSSSTDLG.....QKYIAKLSLNSLYGKFGQKE 661: AVQISAMISSYARTSINPF
pCLKI 781: ..VKCNPNNVT.....EKNIAKLILNSLIGRFGMNI 883: SITTAAAVLSYARIHMAQI
pLMP1 1102: ..LKTEAEMNKDD..ISRLIYKLILNSLSGRWGLRD 1208: AVQLSAAISAYGRISMSKV
pAI2 936: ...SSYTP.DDPM....YFIAKLLMNSLYGRFGMDP 1013: SVPISAAIAAYSRMIMSEY
pAL2-1 782: ...KANSEGSEK.....LITKFLLNSLLGRFGMSI 883: AVSISAAVTAYARIFMAQT
PRD1 322: ..KRDAAKKACDL..FHNIFYKLILNSSYGKFAQNP 398: NVAVAASITGAARSVLLRA
phi29 368: ..YIKTTSEGA.....IKQLAKLMLNSLYGKFASNP 425: YTPMGVFITAWARYTTITA
Adeno2 683: ....RDKNQLT.....RSIAKLLSNALYGSFATKL 827: PSHIASFVLAWIRAFVSEW

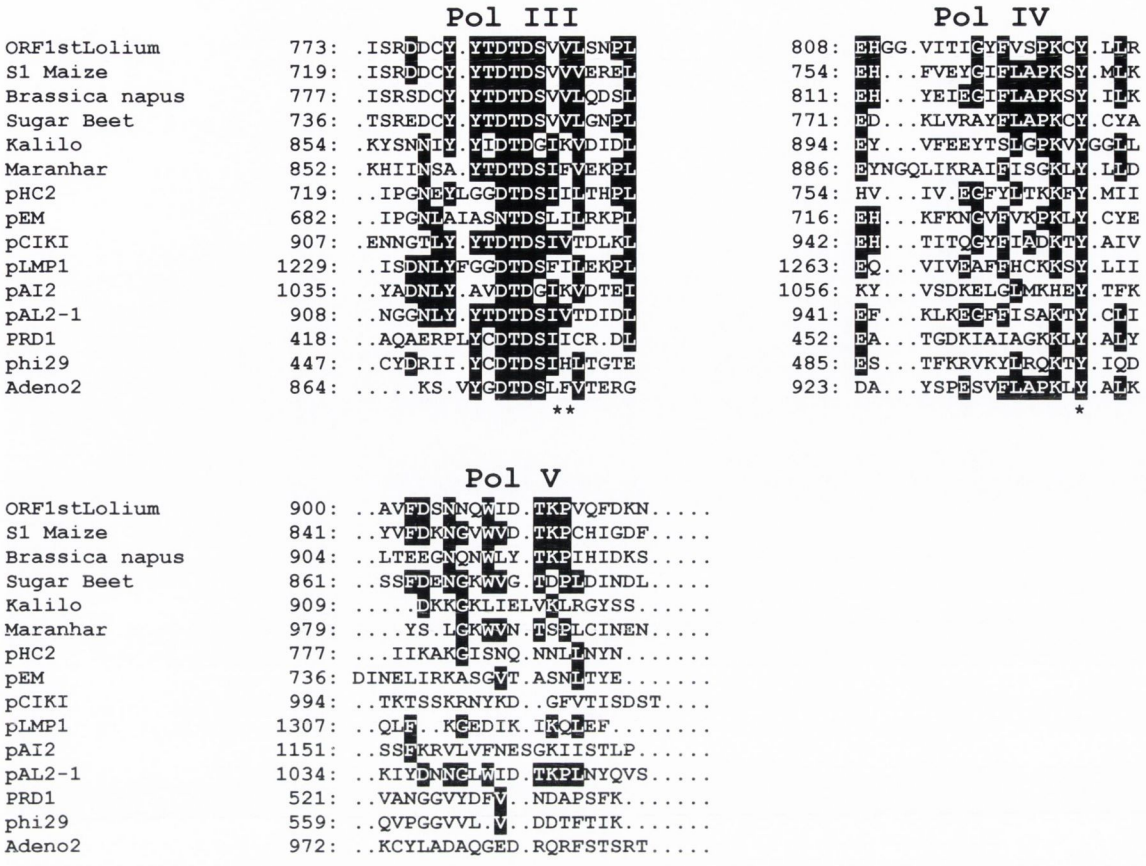
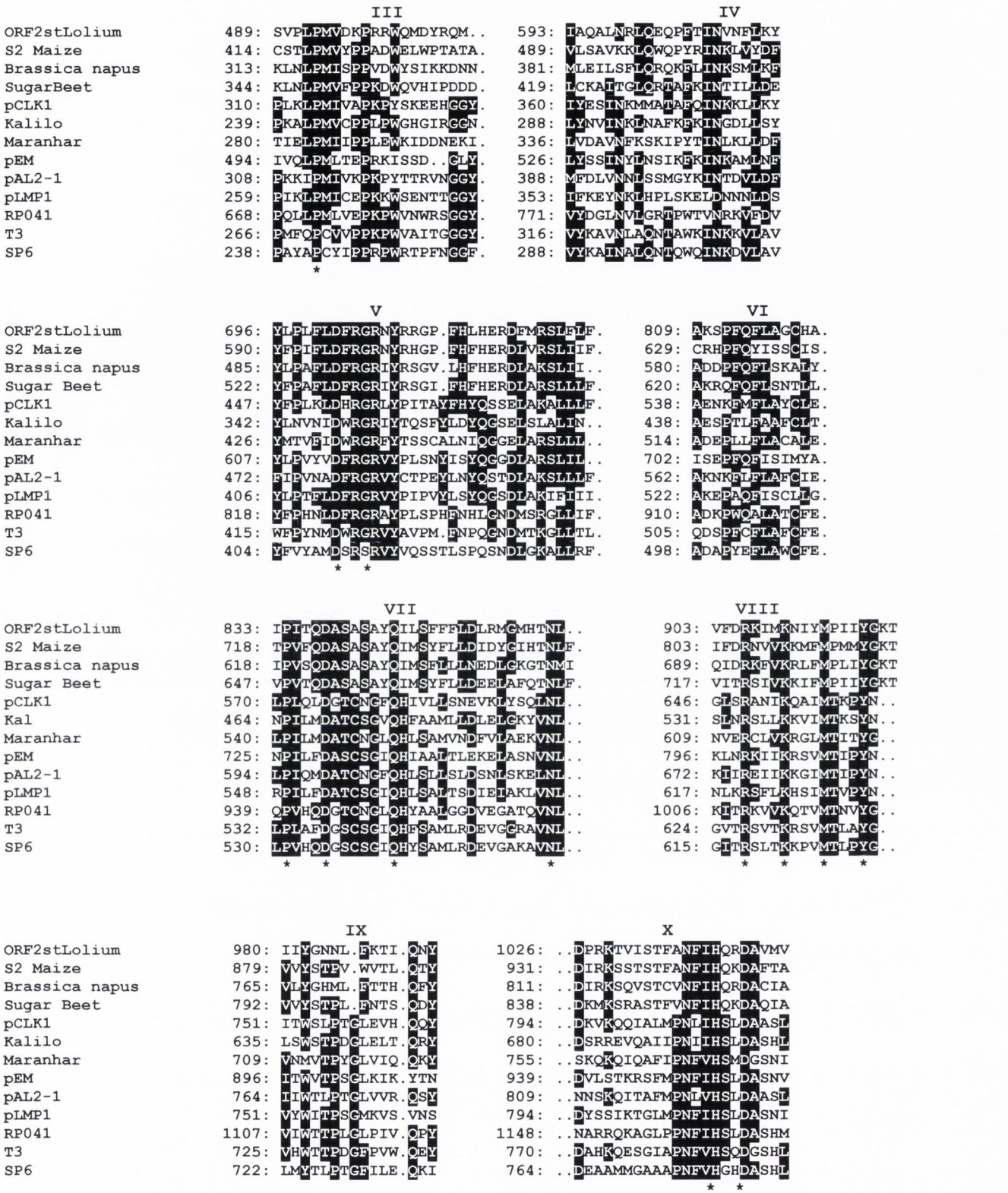


Figure 4.15 Conserved DNA Polymerase Domains in ORF1 for *L. perenne*

A multiple alignment of the putative ORF1 DNA Polymerase protein in the mitochondrial genome of *Lolium perenne* CMS9B290 with several members of the DNA B Polymerase family consisting of bacteriophage, viruses, mitochondrial plasmids in plant and fungi. S1 plasmid in *Zea mays* (Accession no. X02451), 11.6 kb plasmid in *Brassica napus* (Accession no. AB073400), 10.4 kb plasmid in *Beta vulgaris var.maritime* (Accession no. Y10854), Kalilio plasmid in *Neurospora crassa* (Accession no. X52106) Maranhar plasmid in *Neurospora intermedia* (Accession no. X55361), pHC2 plasmid in *Hebeloma circinans* (Accession no. CAA72280), pEM plasmid in *Agaricus bitorquis* (Accession no.CAA44800), pCLK1 plasmid in *Claviceps purpurea* (Accession no. P22373), pLMP1 plasmid in *Pleurotus ostreatus* (Accession no.AF126285), pAI2 plasmid in *Ascobolus immerus* (Accession no. S05342), pAL2-1 in plasmid *Podospora anserina* (Accession no. QQ1529), bacteriophage PRD1 DNA Polymerase (Accession no. M69077), phi29 bacteriophage DNA Polymerase in *Bacillus subtilus* (Accession no. P03680), Adenoviruse Adeno2 DNA Polymerase (Accession no.AAA92206). DNA Polymerases with an identity greater than four amino acid residues are shaded in black. The aesterix symbol * represents amino acid residue that shows 100% identity in all of the systems chosen. The symbol † represents the conserved Asp-D residue in the TRR-1 domain.


III
IV
ORF2stLolium 489: SVPLPMVDKPRRWQMDYRQM.. 593: IAQALNRLQEQPFTINVNFLKY
S2 Maize 414: CSTLPMVYPPADWELWPTATA. 489: VLSAVKKLQWQPYRINKLVYDF
Brassica napus 313: KLNLPMISPPVDWYSIKKDNN. 381: MLEILSFLQRQKFLINKSMLKF
SugarBeet 344: KLNLPMVFPPKDWQVHIPDDD. 419: LCKAITGLQRTAFKINTILLDE
pCLK1 310: PLKLPMIVAPKPYSKEEHGGY. 360: IYESINKMMATAFQINKKLLKY
Kalilo 239: PKALPMVCPPLPWGHGIRGGN. 288: LYNVINKLNAFKFKINGDLLSY
Maranhar 280: TIELPMIIPPLEWKIDDNEKI. 336: LVDAVNFKSKIPYTINLKLLDF
pEM 494: IVQLPMLTEPRKISSD..GLY. 526: LYSSINYLNSIKFKINKAMLNF
pAL2-1 308: PKKIPMIVKPKPYTTRVNGGY. 388: MFDLVNNLSSMGYKINTDVLDF
pLMP1 259: PIKLPMICEPKKWSENTTGGY. 353: IFKEYNKLHPLSKELDNNNLDS
RP041 668: PQLLPMLVEPKPWVNWRSGGY. 771: VYDGLNVLGRTPWTVNRKVFDV
T3 266: PMFQPCVVPPKPWVAITGGGY. 316: VYKAVNLAQNTAWKINKKVLAV
SP6 238: PAYAPCYIPPRPWRTPFNGGF. 288: VYKAINALQNTQWQINKDVLAV
V
VI
ORF2stLolium 696: YLPLFLDFRGRNYRRGP.FHLHERDFMRSLFLF. 809: AKSPFQFLAGCHA.
S2 Maize 590: YFPIFLDFRGRNYRHGP.FHFHERDLVRSLIIF. 629: CRHPFQYISSCIS.
Brassica napus 485: YLPAFLDFRGRIYRSGV.LHFHERDLAKSLII.. 580: ADDPFQFLSKALY.
Sugar Beet 522: YFPAFLDFRGRIYRSGI.FHFHERDLARSLLLF. 620: AKRQFQFLSNTLL.
pCLK1 447: YFPLKLDHRGRLYPITAYFHYQSSELAKALLLF. 538: AENKFMFLAYCLE.
Kalilo 342: YLNVNIDWRGRIYTQSFYLDYQQSELSLALIN.. 438: AESPTLFAAFCLT.
Maranhar 426: YMTVFIDWRGRFYTSSCALNIQGGELARSLLL.. 514: ADEPLLFLACALE.
pEM 607: YLPVYVDFRGRVYPLSNYISYQGGDLARSLIL.. 702: ISEPFQFISIMYA.
pAL2-1 472: FIPVNADFRGRVYCTPEYLNYQSTDLAKSLLLF. 562: AKNKFLFLAFCIE.
pLMP1 406: YLPTFLDFRGRVYPIPVYLSYQQGSDLAKIFIII. 522: AKEPAQFISCLLG.
RP041 818: YFPHNLDFRGRAYPLSPHFNHLGNDMSRGLLIF. 910: ADKPWQALATCFE.
T3 415: WFPYNMDWRGRVYAVPM.FNPQGNDMTKGLLTL. 505: QDSPFCFLAFCFE.
SP6 404: YFVYAMDSRSRVYVQSSTLSPQSNDLGKALLRF. 498: ADAPYEFLAWCFE.
VII
VIII
ORF2stLolium 833: IPITQDASASAYQILSFFFLDLRMGMHTNL.. 903: VFDRKIMKNIYMPIIYGKT
S2 Maize 718: TPVFQDASASAYQIMSYFLLDIDYGIHTNLF. 803: IFDRNVVKKMFMPMMYGKT
Brassica napus 618: IPVSQDASASAYQIMSFLLLNEDLGKGTNMI 689: QIDRKFVKRLFMPLIYGKT
Sugar Beet 647: VPVTQDASASAYQIMSYFLLDEELAFQTNLF. 717: VITRSIVKKIFMPIIYGKT
pCLK1 570: LPLQLDGTCNGFQHIVLLSNEVKLYSQLNL.. 646: GLSRANIKQAIMTKPYN..
Kal 464: NPILMDATCSGVQHFAAMLLDLELGKYVNL.. 531: SLNRSLLKKVIMTKSYN..
Maranhar 540: LPILMDATCNGLQHLSAMVNDFVLAEKVNL.. 609: NVERCLVKRGLMTITYG..
pEM 725: NPILFDASCSGIQHIAALTLEKELASNVNL.. 796: KLNRKIIKRSVMTIPYN..
pAL2-1 594: LPIQMDATCNGFQHLSLLSLDSNLSKELNL.. 672: KLIRELIKKGIMTIPYN..
pLMP1 548: RPILFDATCSGIQHLSALTSDIELAKLVNL.. 617: NLKRSFLKHSIMTVPYN..
RP041 939: QPVHQDGTCNGLQHYAALGGDVEGATQVNL.. 1006: KITRKVVKQTVMTNVYG..
T3 532: LPLAFDGSCSGIQHFSAMLRDEVGGRAVNL.. 624: GVTRSVTKRSVMTLAYG.
SP6 530: LPVHQDGSCSGIQHYSAMLRDEVGAKAVNL.. 615: GITRSLTKKPVMTLPYG..
IX
X
ORF2stLolium 980: IIYGNNL.EKTI.QNY 1026: ..DPRKTVISTFANFIHQRDAVMV
S2 Maize 879: VVYSTPV.WVTL.QTY 931: ..DIRKSSTSTFANFIHQKDAFTA
Brassica napus 765: VLYGHML.ETTH.QFY 811: ..DIRKSQVSTCVNFIHQRDACIA
Sugar Beet 792: VVYSTPL.ENTS.QDY 838: ..DKMKSRASTFVNFIHQKDAQIA
pCLK1 751: ITWSLPTGLEVH.QQY 794: ..DKVKQQIALMPNLIHSLDAASL
Kalilo 635: LSWSTPDGLELT.QRY 680: ..DSRREVQAIIPNIIHSLDASHL
Maranhar 709: VNMVTPYGLVIQ.QKY 755: ..SKQKQIQAFIPNFVHSMDGSNI
pEM 896: ITWVTPSGLKIK.YTN 939: ..DVLSTKRSFMPNFIHSLDASNV
pAL2-1 764: IIWTLPTGLVVR.QSY 809: ..NNSKQITAFMPNLVHSLDAASL
pLMP1 751: VYWITPSGMKVS.VNS 794: ..DYSSIKTGLMPNFIHSLDASNI
RP041 1107: VIWTTPLGLPIV.QPY 1148: ..NARRQKAGLPPNFIHSLDASHM
T3 725: VHWTTPDGFPVW.QEY 770: ..DAHKQESGIAPNFVHSQDGSHL
SP6 722: LMYTLPTGFILE.QKI 764: ..DEAAMMGAAAPNFVHGHDASHL

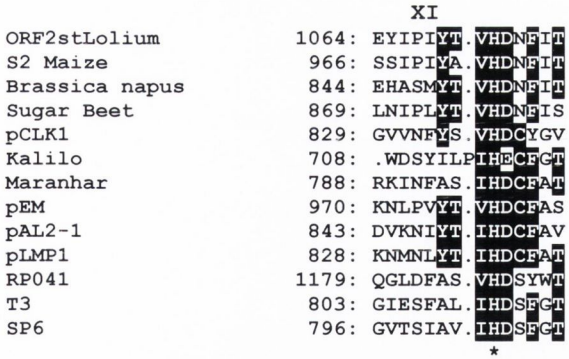


Figure 4.16 Conserved RNA Polymerase Domains in ORF2 for *L. perenne*

Comparison of the putative RNA polymerase ORF2 in the mitochondrial genome of *Lolium perenne* CMS9B290 with previously identifiable conserved domains of a single subunit RNA Polymerases (DNA directed RNA Polymerase) encoded by several mitochondrial plasmids in fungi and plant, phage and in a yeast mitochondrial encoded genome. Highlighted by a black shade are five or more conserved residues. The linear plasmids examined in fungi were; pCLK1 *Claviceps purpurea* (Accession no. P22372), Kalilo *Neurospora crassa* (Accession no. X52106), Maranhar *Neurospora intermedia* (Accession no. X55361), pEM *Agaricus bisporus* (Accession no.AAD11785), pAL2-1 *Podospora anserinia* (Accession no.S26945), pLMP1 *Pleurotus ostreatus* (Accession no. AF126285), linear plasmids in plant were; S2 *Zea mays* (Accession no. J01426), 11.64kb plasmid *Brassica napus* (Accession no. AB073400), 10.4kb plasmid *Beta vulgaris var. maritima* (Accession no. Y10854), mitochondrial RNA Polymerase in yeast; RPO41 *Saccharomyces cerevisiae* (Accession no.NP116617), bacteriophage SP6 RNA Polymerase (Accession no. CAA68288) and phage T3 RNA Polymerase (Accession no. RNBPT3). RNA Polymerases with an identity greater than four amino acid residues are shaded in black. The aesterix symbol * represents amino acid residue that shows 100% identity in all of the systems chosen.

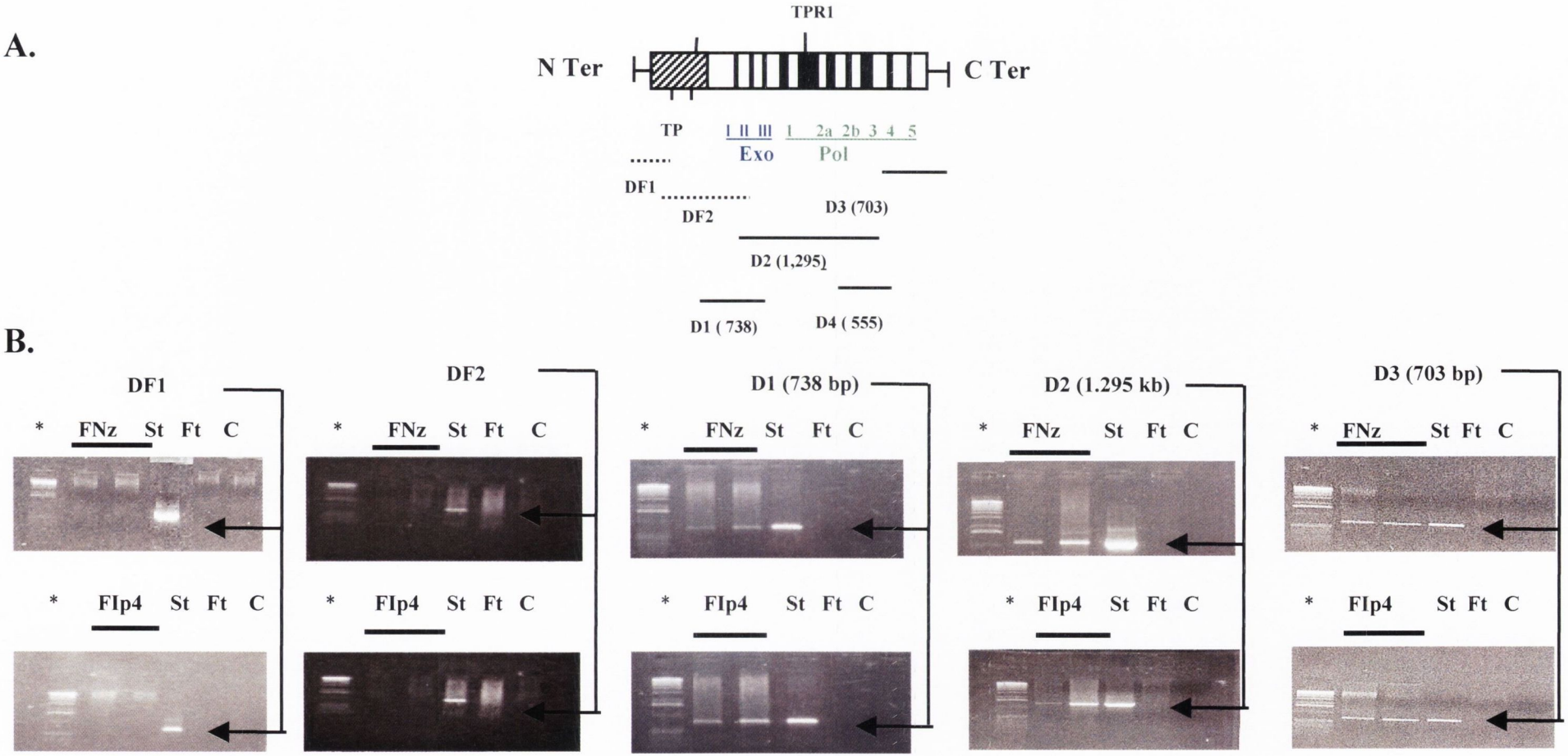
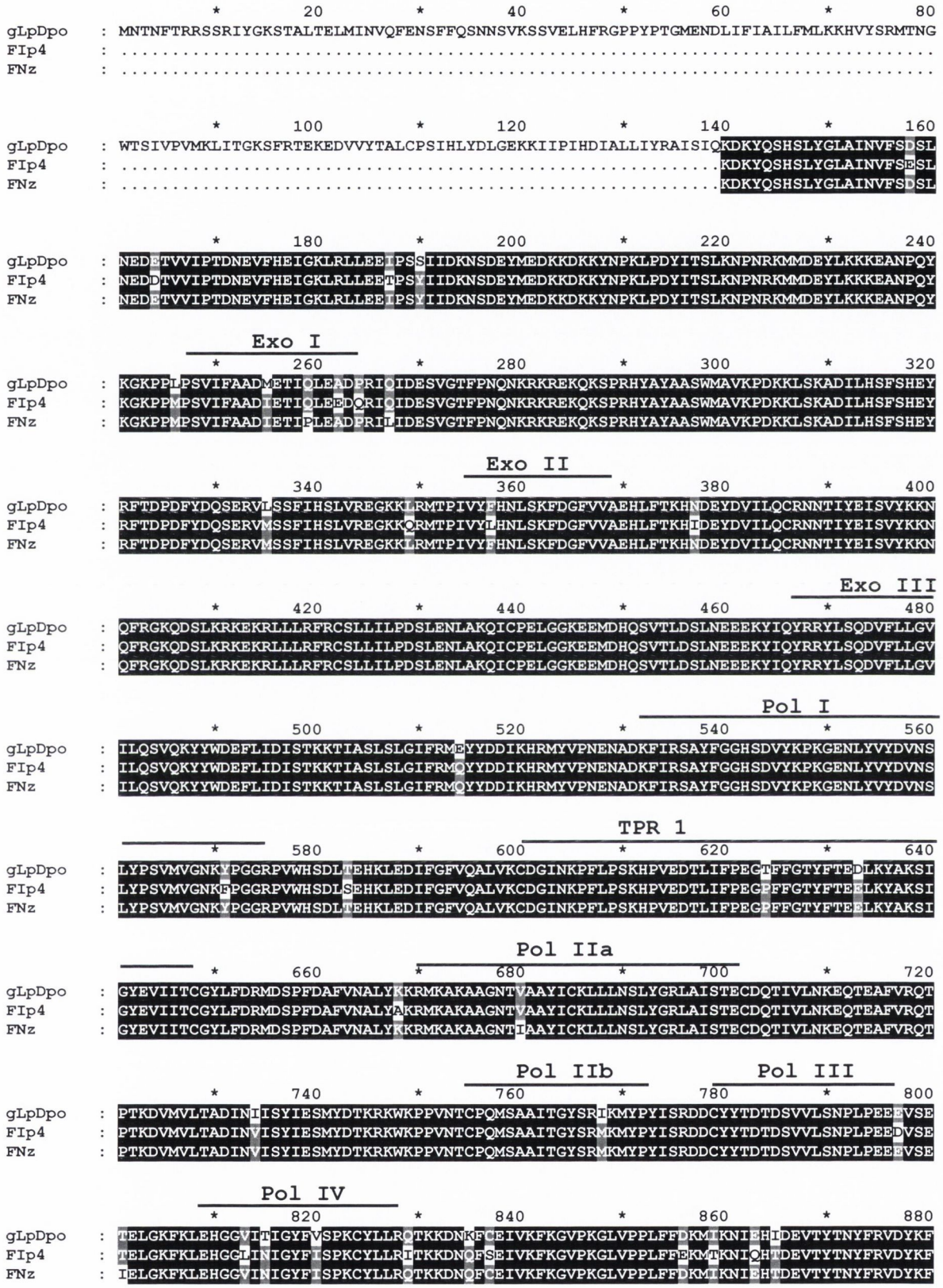


Figure 4.17 PCR products amplified for Dpo ORF1 sequence in sterile and fertile lines.

A. Is a map of some of the primers used to amplify conserved regions of ORF1 in the sterile line-St, the fertile lines FNz and FIp4.No PCR product was seen in the fertile wild type lineFt. The conserved domains are dark boxes, the hatched box is a putative terminal protein TP. N-Ter is upstream from ATG start codon and C-Ter is downstream sequence from the TAA stop codon of ORF1. **B.** The sizes of the amplified PCRfragments are highlighted in bracketts. There are four PCR fragments; D1 is 738 bp in size, D2 is 1.295 kb in size, D3 is 703 bp in size, D4 is 555 bp in size. DF1 and DF2 fragments are only amplified in the sterile line-St. The lane * refers to the 1 kb marker.


gLpDpo
FIp4
FNz
Exo I
Exo II
Exo III
Pol I
TPR 1
Pol IIa
Pol IIb
Pol III
Pol IV

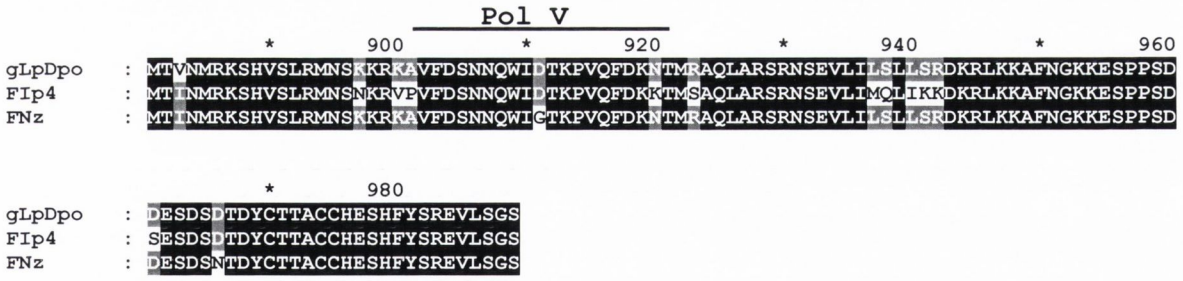


Figure 4.18 Alignment of a DNA Polymerase amino acid sequence in *L. perenne*

The ORF1 (gLpDpo) amino acid sequence in Sterile line of *L. perenne* CMS9B290 compared with Dpol like sequence in fertile lines of *L. perenne* FIp4 and FNz (that were subjected to an endophytic infection). The dark shaded box refers to identical amino acids in the three lines, the light shaded box refers to identical amino acids in two lines only, and non-shaded box are non-conserved amino acids. Roman numerals refer to the different conserved domains.

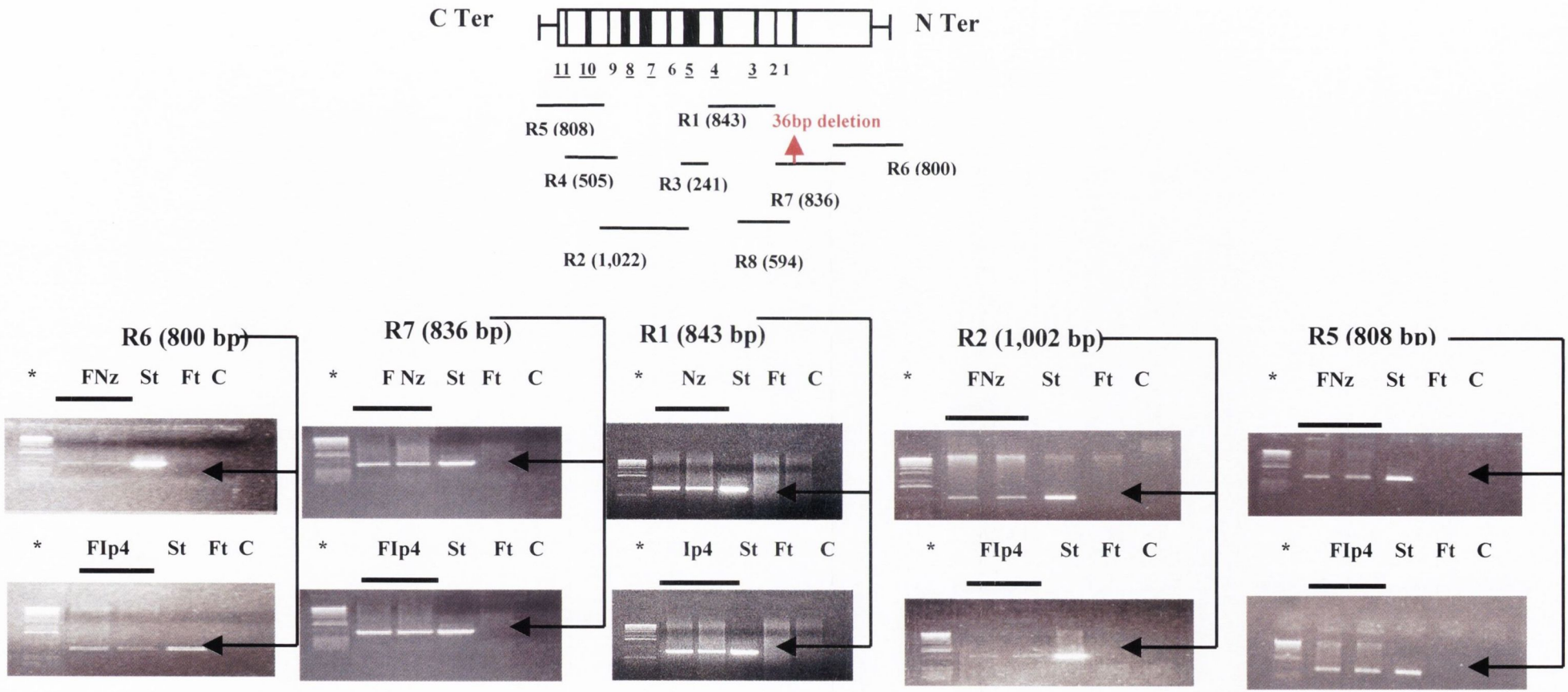
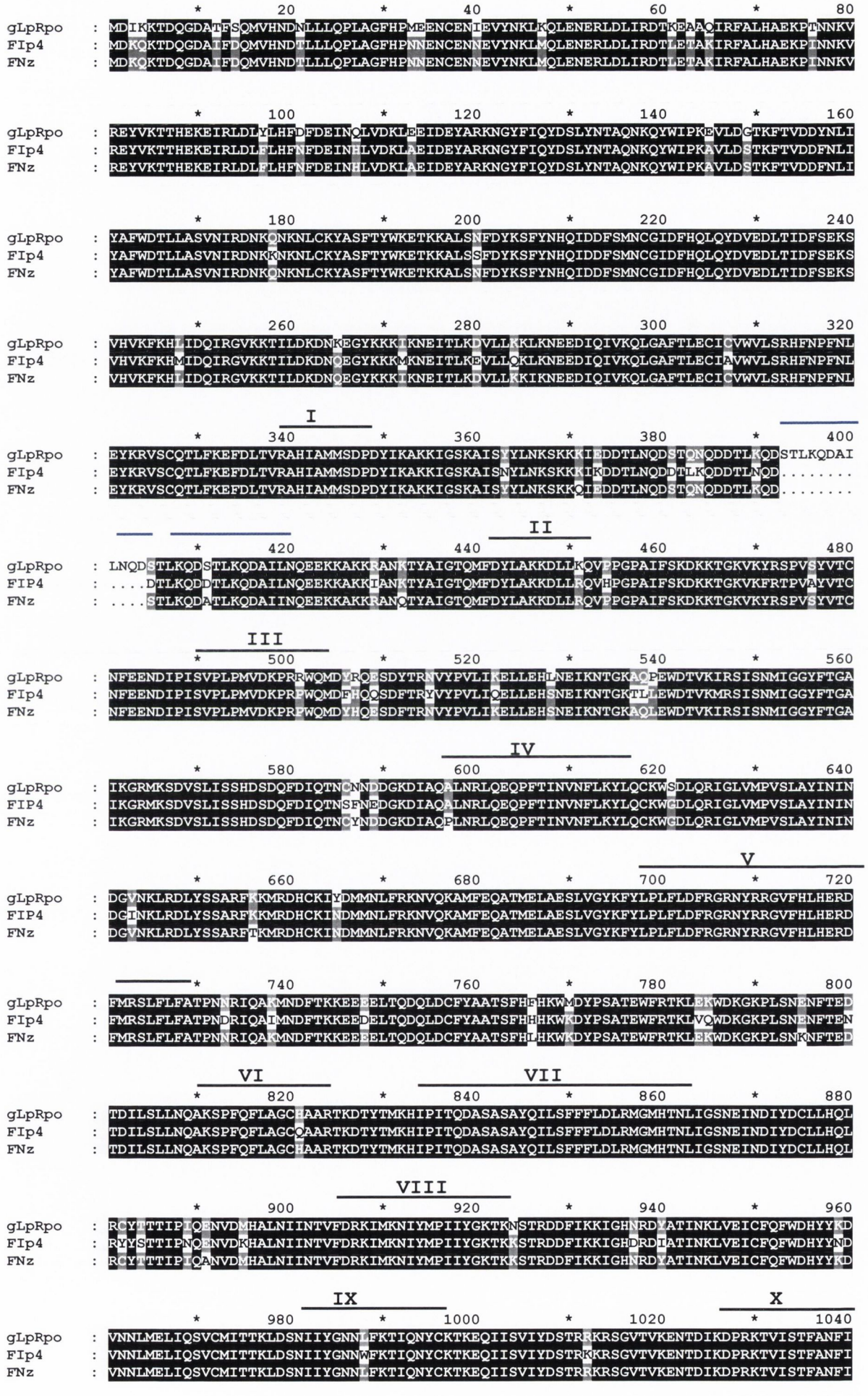


Figure 4.19 PCR products amplified for Rpo ORF2 sequence in sterile and fertile lines.

A. Is a map of some of the primers used to amplify conserved regions of ORF2 in the sterile line-St, the fertile lines FNz and FIp4.No PCR product was seen in the fertile wild type lineFt.. N-Ter is upstream from ATG start codon and C-Ter is downstream sequence from the TAA stop codon of ORF2. **B** The sizes of the amplified PCR fragments are highlighted in bracketts. There are four PCR fragments; R1 is 843 bp in size, R2 is1,002 bp in size, R5 is 808 bp in size, R6 is 800 bp in size, and R7 is 836 bp in size. R3 are fragments overlaps for R1, R2; R8 are fragments overlaps for R1, R7. The lane * refers to the 1 kb marker.


gLpRpo
FIp4
FNz
I
II
III
IV
V
VI
VII
VIII
IX
X

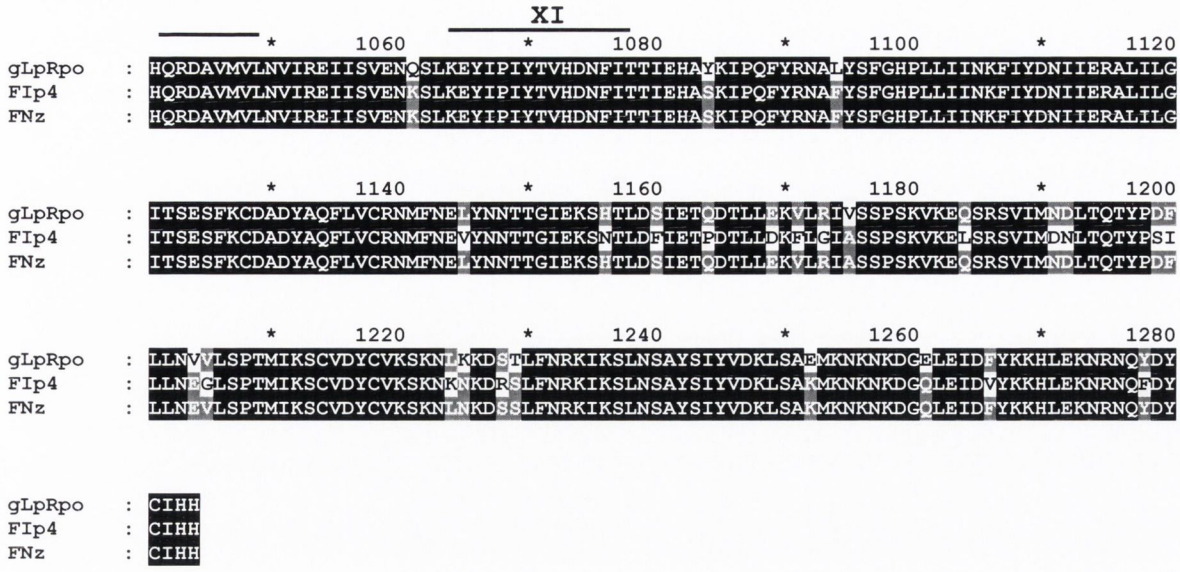


Figure 4.20 Alignment of RNA Polymerase amino acid sequence in *L. perenne*

The ORF2 (gLpRpo) amino acid sequence in Sterile line of *L. perenne* CMS9B290 compared Rpol like sequence in two fertile lines of *L. perenne* FIp4 and FNz (subjected to an endophytic infection). The dark shaded box refers to identical amino acids in all three lines, the light shaded box refers to identical amino acids in two lines only, and non-shaded box are different amino acids. Roman numerals refer to the different domains, blue line ——— refers to deleted/repeated amino acid sequence.

A.

```
gLpRpo        CATATTGCCATGATGTCGGATCCAGATTATATTAAGGCAAAGAAAATAGGATCAAAAGCG 1080
FIp4          CATATTGCCATGATGTCGGATCCAGATTATATTAAGGCAAAGAAAATAGGATCAAAAGCG 1080
FNz           CATATTGCCATGATGTCGGATCCAGATTATATTAAGGCAAAGAAAATAGGATCAAAAGCG 1080
              ************************************************************
                                                                (i)
                                                          D  T  L  N  Q  D
gLpRpo        ATTTCCTATTATCTTAATAAAAGCAAAAAGAAAATCGAAGATGATACTCTAAATCAAGAT 1140
FIp4          ATTTCCAATTATCTTAATAAAAGCAAAAAGAAAATCAAAGATGATACTCTAAATCAAGAT 1140
FNz           ATTTCCTATTATCTTAATAAAAGCAAAAAGCAAATCGAAGATGATACTCTAAATCAAGAT 1140
              ****** *********************** ***** ***********************
                     (ii)                (iii)                (iv)      (v)
               S  T  Q  N  Q  D  D  T  L  K  Q  D  S  T  L  K  Q  D  A  I
gLpRpo        TCTACTCAAAATCAAGATGATACTCTAAAACAAGATTCTACTCTAAAACAAGATGCTATT 1200
FIp4          GATACTCTAAAACAAGATGATACTCTAAATCAAGAT------------------------ 1198
FNz           TCTACTCAAAATCAAGATGATACTCTAAAACAAGAT------------------------ 1198
                ***** *** ***************** ******
                                   (vi)           (vii)            (viii)
              L  N  Q  D  S  T  L  K  Q  D  S  T  L  K  Q  D  A  I  L  N
gLpRpo        CTAAATCAAGATTCTACTCTAAAACAAGATTCTACTCTAAAACAAGATGCTATTCTAAAT 1260
FIp4          ------------GATACTCTAAAACAAGATGATACTCTAAAACAAGATGCTATTCTAAAT 1224
FNz           ------------TCTACTCTAAAACAAGATGCTACTCTAAAACAAGATGCTATTATAAAT 1224
                             ***************** ********************** *****
              Q  E
gLpRpo        CAAGAAGAAAAGAAAGCAAAAAAGAGAGCAAATAAAACTTATGCTATAGGGACACAGATG 1320
FIp4          CAAGAAGAAAAGAAAGCAAAAAAGATAGCAAATAAAACTTATGCTATAGGGACACAGATG 1284
FNz           CAAGAAGAAAAGAAAGCAAAAAAGAGAGCAAATCAAACTTATGCTATAGGGACACAGATG 1284
              ************************* ******* *************************
```

B.

```
gLpRpo (i)       TACTCTAAATCAAGAT
gLpRpo (ii)      TACTCAAAATCAAGAT
gLpRpo (iii)     TACTCTAAAACAAGAT
gLpRpo (iv)      TACTCTAAAACAAGAT
gLpRpo (v)       TATTCTAAATCAAGAT
gLpRpo (vi)      TACTCTAAAACAAGAT
gLpRpo (vii)     TACTCTAAAACAAGAT
gLpRpo (viii)    TATTCTAAATCAAGAA
                 ** ** *** *****
```

Figure 4.21 Alignment of Nucleotide sequence for repeated region in Rpo

A. Is a repetitive region in gLpRpo, FIp4 and FNz showing a duplication of the nucleotide acid sequence as underlined in the regions (i), (ii), (iii), (vi), (vii), (viii). Also underlined are the regions (iv) and (v) representing an additional duplication present only in gLpRpo, but absent in FIp4 and FNz. **B.** Is a multiple alignment (ClustalW) of the nucleotide sequence for each of the duplicated regions present in gLpRpo. The aesterick symbol * corresponds to identical nucleotides for each of the alignments.

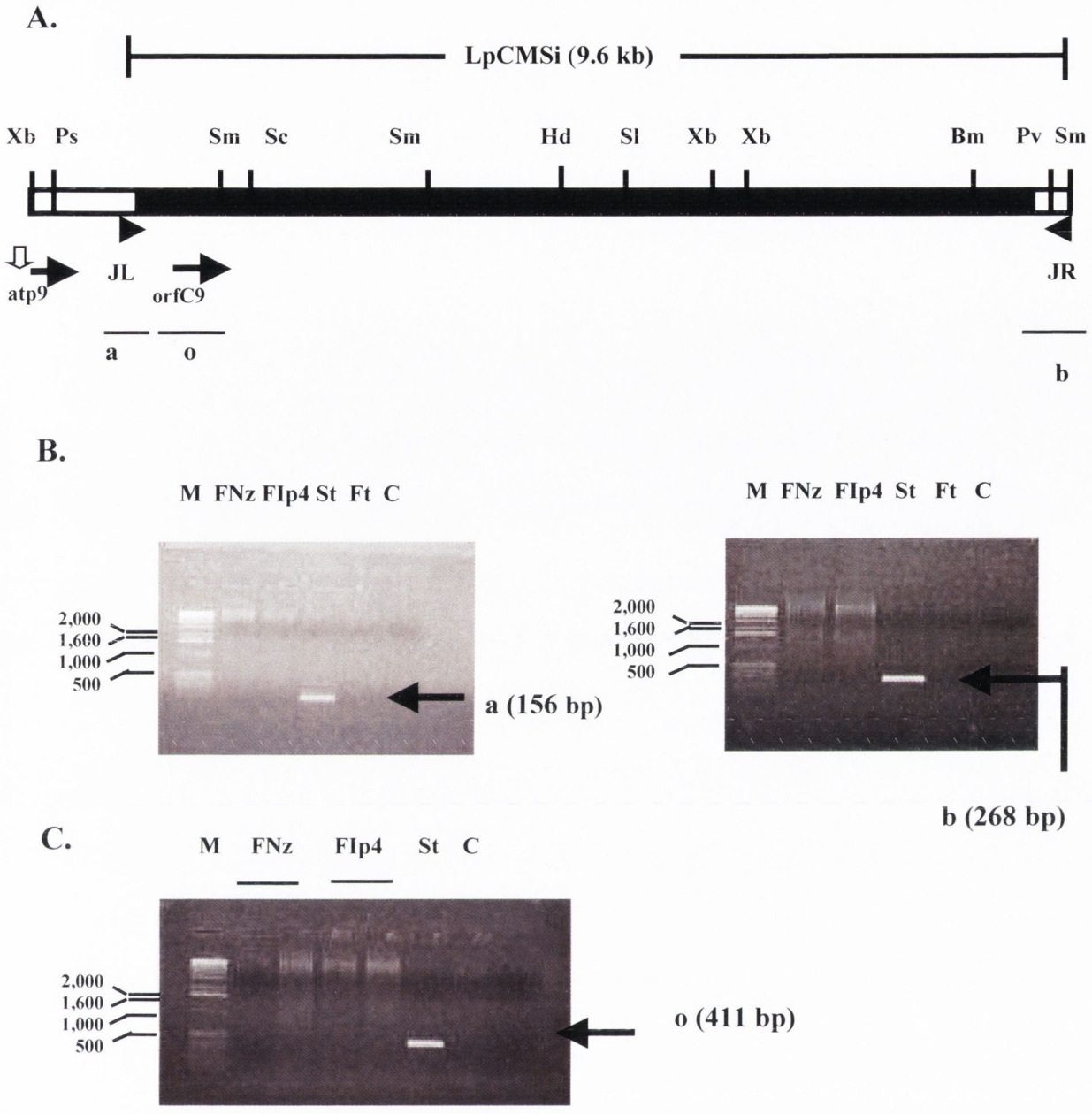


Figure 4.22 PCR Confirmation in the Integration of the 9.6 kb CMS associated region in *L.perenne*.

A. Map of LpCMSi-9.6 kb region showing position of junction fragments a, b and chimeric orfC9 fragment o. **B.** PCR amplification (30 cycles) of the segments a (156bp) and b (268bp) spanning the junction sites JL, JR. **C.** PCR amplification of the fragment o (411bp) spanning the chimeric *orfC9* region in the mitochondria of the Sterile line – S*t*. The negative control (no template) is highlighted by the lane C and 1 kb marker by the lane M. Lanes FNz, FIp4 and Ft refers to fertile lines of *L. perenne* and St refers to the line CMS-St (CMS9B290).

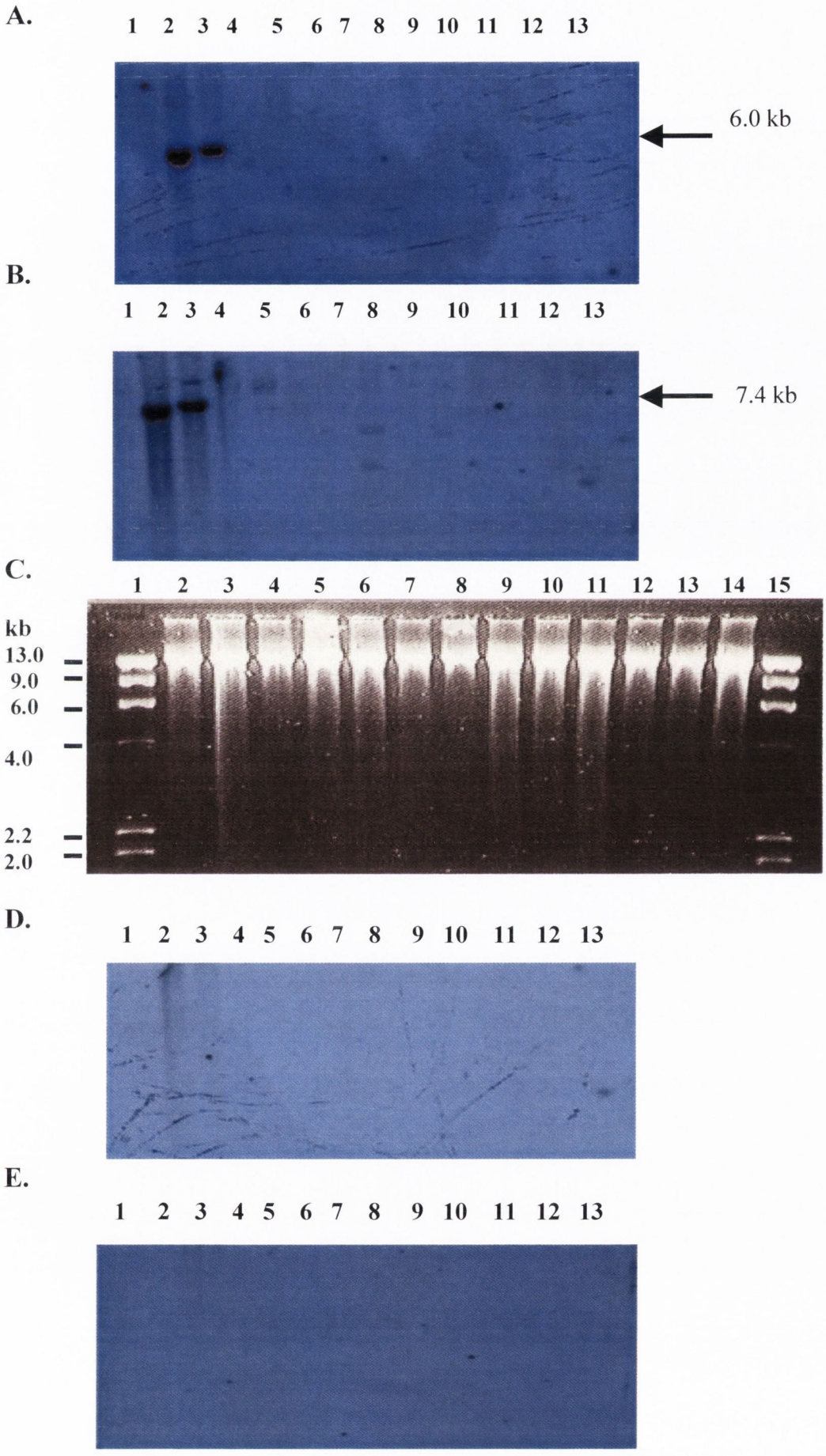

A.
1 2 3 4 5 6 7 8 9 10 11 12 13
6.0 kb
B.
1 2 3 4 5 6 7 8 9 10 11 12 13
7.4 kb
C.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
kb
13.0
9.0
6.0
4.0
2.2
2.0
D.
1 2 3 4 5 6 7 8 9 10 11 12 13
E.
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 4.23 Southern blot of Proteinase K treated and digested total cellular DNA in *L. perenne.*

A. Total DNA digested with *Hind*III probed with Dpo probe (PolI-III) **B.** Total DNA digested with *Hind*III probed with Rpo probe (RII-V). **C.** Total DNA from lines of *L. perenne* digested with proteinase K. Lanes 1 and 15 is the lambda/*Hin*dIII marker. Lane 2 is fertile wild type-Ft, lane 3 and lane 4 is sterile line-St, lane 5 is fertile line-FNz, lane 6 to lane 14 is the fertile Ip lines-Ip2, Ip3, Ip4, Ip4, Ip5, Ip6, Ip7, Ip8, Ip9 respectively. **D.** Proteinase K treated total cellular DNA probed with Dpol probe (PolI-III). **E.** Proteinase K treated total cellular DNA probed with Rpol probe (RII-V).

In all four Southern blots (A, B, D , E) lane 1 is fertile wild type-Ft, lanes 2 and 3 are sterile lines St, lane 4 is fertile line-FNz, lanes 5, 6, 7, 8, 9, 10, 11, 12, 13 are fertile lines Ip2, Ip3, Ip4, Ip4, Ip5, Ip6, Ip7, Ip8, Ip9 respectively.

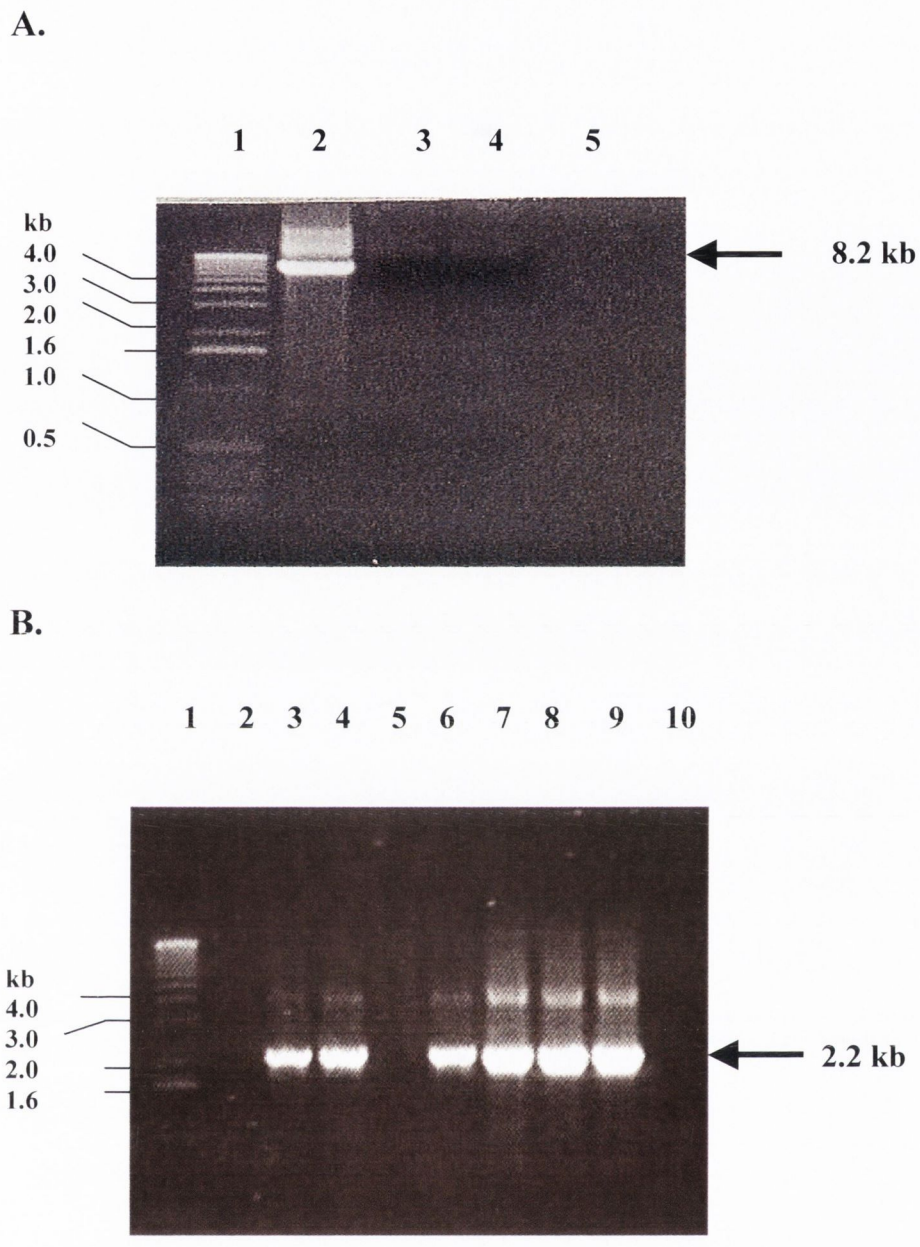


Figure 4.24 PCR amplification of the intergenic region between ORF1 and ORF2.

A. PCR product of 8.2 kb in size amplified in sterile line only after 35 cycles with primers ORF1/2a by long range PCR . Lane 2 is the sterile line (St), lane 3 is the fertile line (FNz), lane 4 is the fertile line (FIp4), lane 5 is the negative control (minus template). **B.** PCR product of 2.2 kb amplified in several lines after 30 cycles with primers ORF1/2b. Lane 2 and 5 is fertile wild type-Ft, lane 3 and 4 are fertile line –FNz, lane 6 is fertile line-FIp4, lane 7 and 8 is the sterile line CMS-St, lane 9 is fertile restored line-Rst1. Lane 10 is the negative control (minus template). Lane 1 is the 1 kb marker.

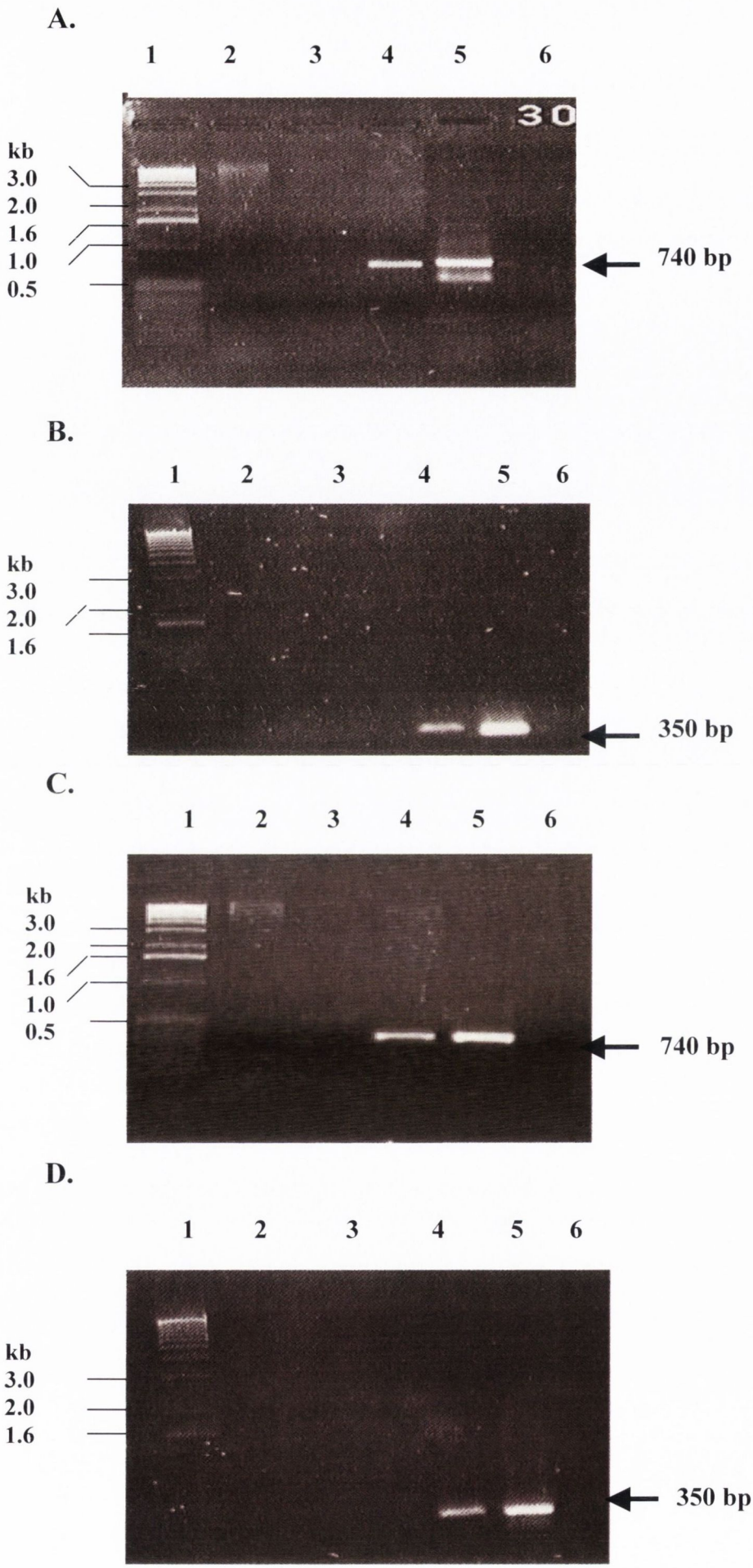

A.
1 2 3 4 5 6
kb
3.0
2.0
1.6
1.0
0.5
740 bp
B.
1 2 3 4 5 6
kb
3.0
2.0
1.6
350 bp
C.
1 2 3 4 5 6
kb
3.0
2.0
1.6
1.0
0.5
740 bp
D.
1 2 3 4 5 6
kb
3.0
2.0
1.6
350 bp

Figure 4.25 PCR amplification of the fungal associated genes *pyr4* and *ß-tubulin* in *L. perenne*.

A. PCR fragment of 740 bp in size, that has been amplified for a large portion in the *pyr4* gene. Lane 1 is the 1 kb marker, lane 2 and 3 fertile lines FIp2 and FIp4 respectively, lane 4 is the fertile line-FNz, lane 5 is the positive control, the ergot fungal line-Fyal, lane 6 is the negative control (minus template). **B.** PCR fragment of 350 bp in size, that has been amplified for a portion of the *ß-tubulin* gene. Lane 1 is the 1kb marker, lane 2 is fertile line-Ft, lane 3 is the sterile line-St, lane 4 is the fertile line-FNz, lane 5 is the positive control , the ergot fungal line-Fyal, lane 6 is the negative control (minus template). **C.** PCR fragment of 740 bp in size, that has been amplified for a large portion in the *pyr4* gene. Lane 1 is the 1kb marker, lane 2 and 3 fertile lines FIp2 and FIp4 respectively, lane 4 is the fertile line-FNz, lane 5 is the positive control, the ergot fungal line-Fyal, lane 6 is the negative control (minus template). **D.** PCR fragment of 350 bp in size amplified for a portion of the *ß-tubulin* gene. Lane 1 is the 1 kb marker, lane 2 is fertile line-Ft, lane 3 is the sterile line-St, lane 4 is the fertile line-FNz, lane 5 is the positive control, the ergot fungal line-Fyal, lane 6 is the negative control (minus template).

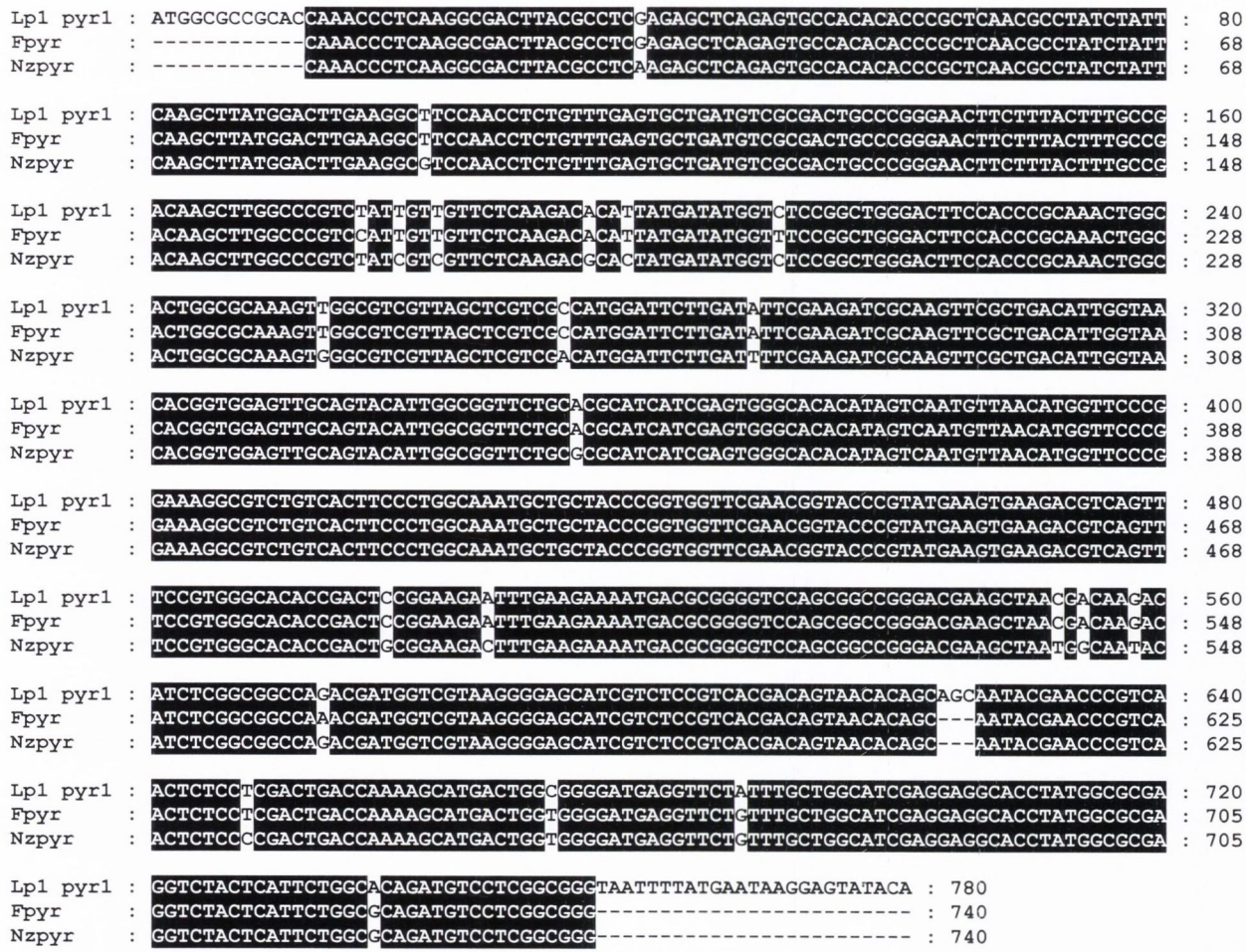


Figure 4.26 Nucleotide Sequence alignment of the *pyr4-1* region in a fertile line of *L. perenne* (FNz) and in fungal endophyte *Epichloe typhia – Acremonium lolli*. The Lp1 *pyr4-1* gene (Acc no. U14564) sequence from an inter-specfic hybrid of endophytes *Epichloe typhina – Acremonium lolli*, and is 97.6% identical with *pyr4-1* like sequence in a fertile line of *L. perenne* (FNz). subjected to an endophyte infection. The sequence Fpyr refers to the *pyr4-1* sequence present in an ergot endophyte (F-yal).

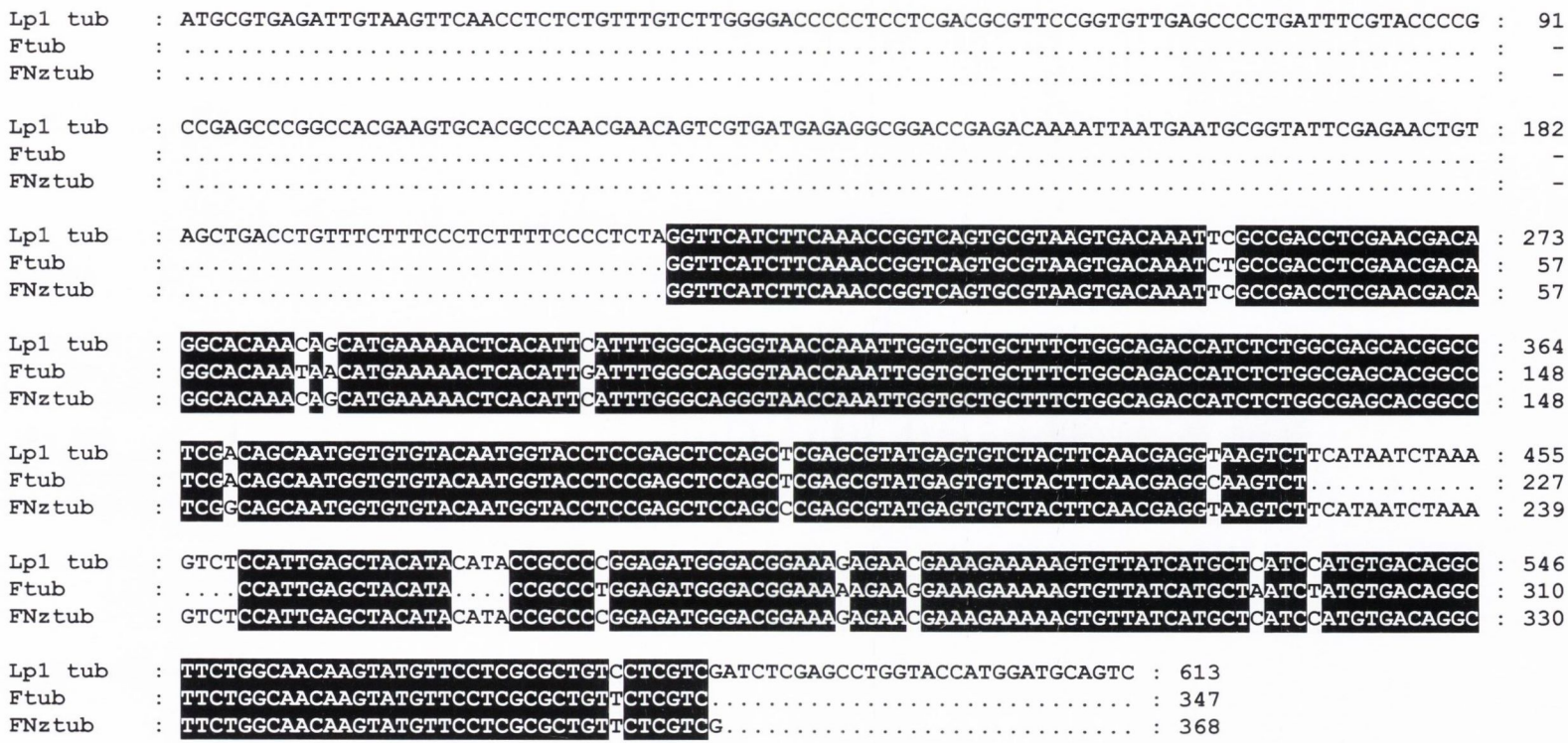
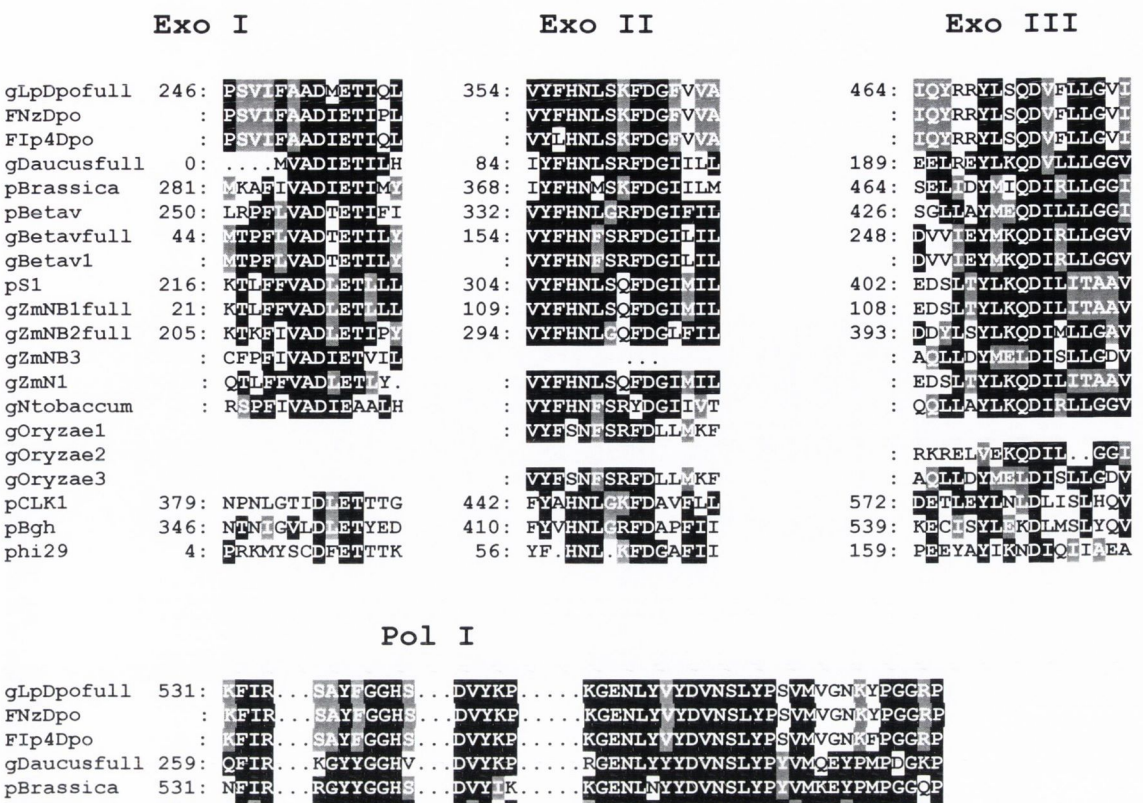
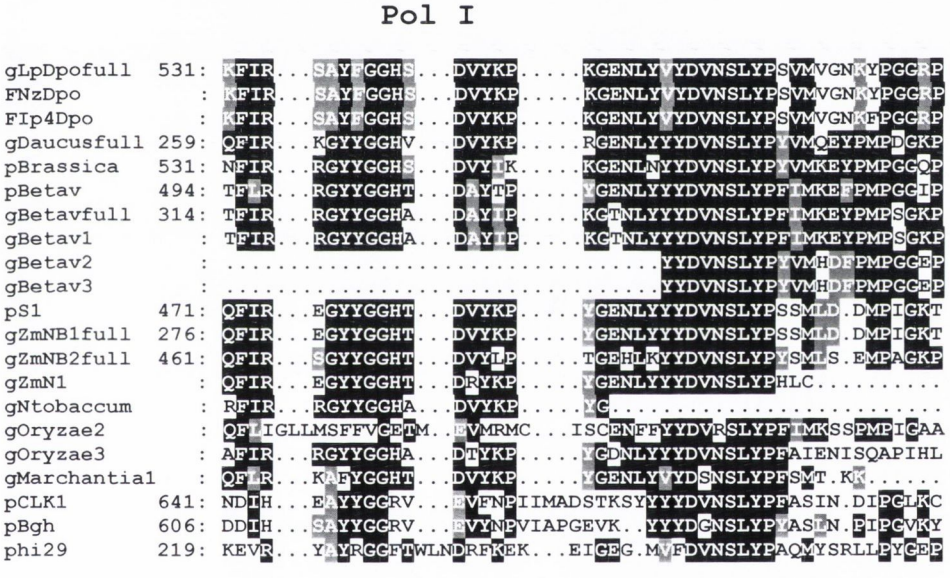
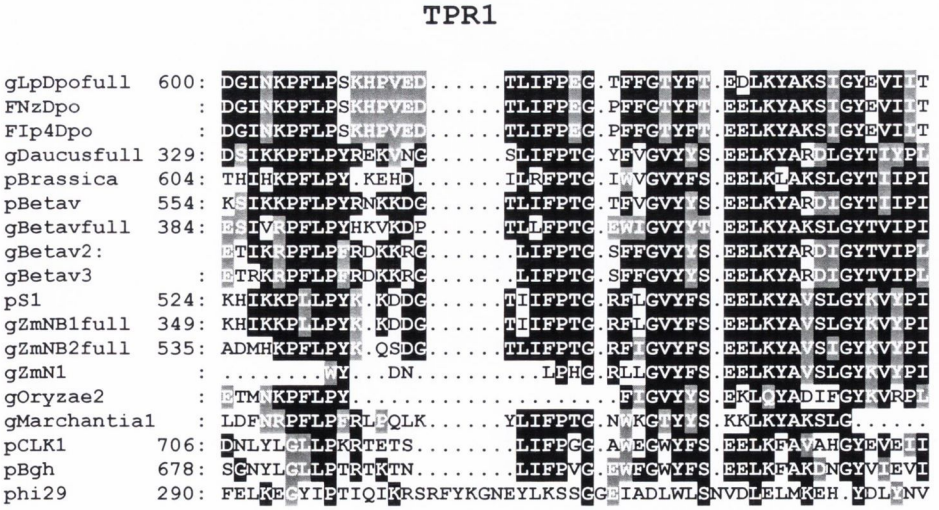
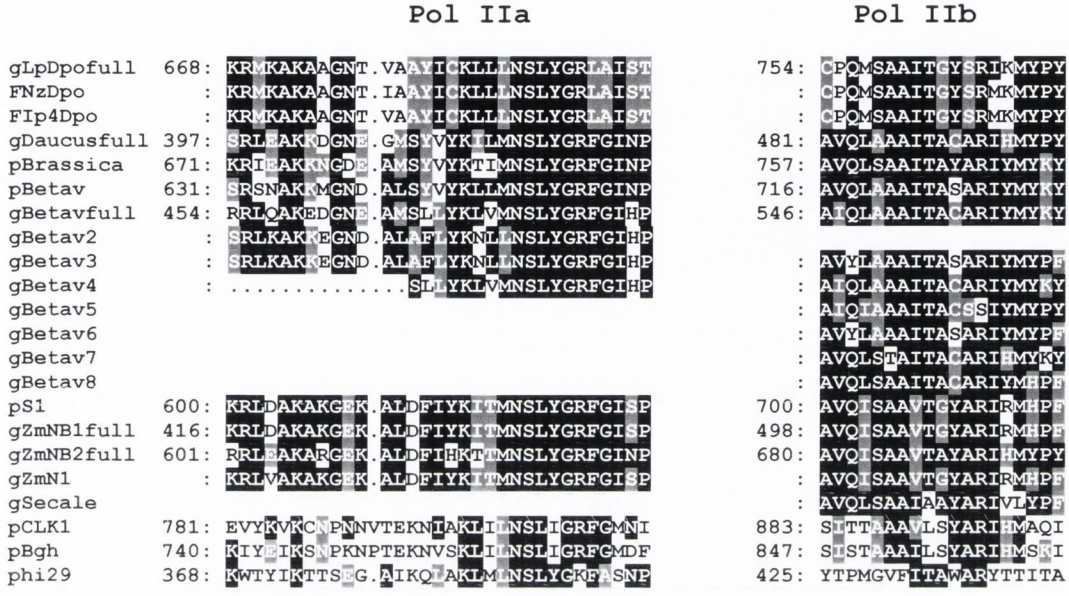


Figure 4.27 Sequences alignment of the *ß*-Tubulin region in a fertile line of *L. perenne* (FNz) and in fungal endophyte *Epichloe typhina – Acremonium lolli.* The Lp1 tub (Acc no. L06956) is the ***ß***-tubulin sequence from an inter-specfic hybrid of endophyte ***Epichloe typhnia – Acremonium lolli,*** and was 99.5% identical with ***ß***-tubulin sequence in fertile line of *L. perenne* (FNz) subjected to endophyte infection. Ftub is a ***ß***-tubulin like sequence present in an ergot endophyte (F-yal).


Exo I
Exo II
Exo III
gLpDpofull 246: PSVIFAADMETIQL 354: VYFHNLSKFDGEVVA 464: IQYRRYLSQDVFLLGVI
FNzDpo : PSVIFAADIETIPL : VYFHNLSKFDGEVVA : IQYRRYLSQDVFLLGVI
FIp4Dpo : PSVIFAADIETIQL : VYLHNLSKFDGEVVA : IQYRRYLSQDVFLLGVI
gDaucusfull 0: ....MVADIETILH 84: IYFHNLSRFDGIILL 189: EELREYLKQDVLLLGGV
pBrassica 281: MKAEIVADIETIMY 368: IYFHNMSKFDGIILM 464: SELIDYMIQDIRLLGGI
pBetav 250: LRPFLVADTETIFI 332: VYFHNLGRFDGIFIL 426: SGLLAYMEQDILLLGGI
gBetavfull 44: MTPFLVADTETILY 154: VYFHNFSRFDGILIL 248: DVVIEYMKQDIRLLGGV
gBetav1 : MTPFLVADTETILY : VYFHNFSRFDGILIL : DVVIEYMKQDIRLLGGV
pS1 216: KTLFFVADLETLLL 304: VYFHNLSQFDGIMIL 402: EDSLIYLKQDILITAAV
gZmNB1full 21: KTLFFVADLETLLL 109: VYFHNLSQFDGIMIL 108: EDSLIYLKQDILITAAV
gZmNB2full 205: KTKGIVADLETIPV 294: VYFHNLGQFDGLFIL 393: DDVLSYLKQDIMLLGAV
gZmNB3 : CFPEIVADIETVIL ... : AQLLDYMELDISLLGDV
gZmN1 : QTLFFVADLETLY. : VYFHNLSQFDGIMIL : EDSLIYLKQDILITAAV
gNtobaccum : RSPEIVADIEAALH : VYFHNFSRYDGIIVT : QQLLAYLKQDIRLLGGV
gOryzae1 : VYFSNFSRFDLLMKF
gOryzae2 : RKRELVEKQDIL..GGI
gOryzae3 : VYFSNFSRFDLLMKF : AQLLDYMELDISLLGDV
pCLK1 379: NPNLGTIDLETTTG 442: FYAHNLGKFDAVELL 572: DETLEYLNLDLISLHQV
pBgh 346: NTNIGVLDLETYED 410: FYVHNLGRFDAPFII 539: KECISYLEKDLMSLYQV
phi29 4: PRKMYSCDFETTTK 56: YF.HNL.KFDGAFII 159: PEEYAMIKNDIQIIEA


Pol I
gLpDpofull 531: KFIR...SAYFGGHS...DVYKP.....KGENLYVYDVNSLYPSVMVGNKYPGGRP
FNzDpo : KFIR...SAYFGGHS...DVYKP.....KGENLYVYDVNSLYPSVMVGNKYPGGRP
FIp4Dpo : KFIR...SAYFGGHS...DVYKP.....KGENLYVYDVNSLYPSVMVGNKFPGGRP
gDaucusfull 259: QFIR...KGYYGGHV...DVYKP.....RGENLYYYDVNSLYPYVMQEYPMPDGKP
pBrassica 531: NFIR...RGYYGGHS...DVYIK.....KGENLNYYDVNSLYPYVMKEYPMPGGQP
pBetav 494: TFLR...RGYYGGHT...DAYTP.....YGENLYYYDVNSLYPFIMKEFPMPGGIP
gBetavfull 314: TFIR...RGYYGGHA...DAYIP.....KGTNLYYYDVNSLYPFIMKEYPMPSGKP
gBetav1 : TFIR...RGYYGGHA...DAYIP.....KGTNLYYYDVNSLYPFIMKEYPMPSGKP
gBetav2 : .....YYDVNSLYPYVMHDFPMPGGEP
gBetav3 : .....YYDVNSLYPYVMHDFPMPGGEP
pS1 471: QFIR...EGYYGGHT...DVYKP.....YGENLYYYDVNSLYPSSMLD.DMPIGKT
gZmNB1full 276: QFIR...EGYYGGHT...DVYKP.....YGENLYYYDVNSLYPSSMLD.DMPIGKT
gZmNB2full 461: QFIR...SGYYGGHT...DVYLP.....TGEHLKYYDVNSLYPYSMLS.EMPAGKP
gZmN1 : QFIR...EGYYGGHT...DRYKP.....YGENLYYYDVNSLYPHLC.....
gNtobaccum : RFIR...RGYYGGHA...DVYKP.....YG.....
gOryzae2 : QFLIGLLMSFFVCETM..EVMRMC...ISCENFFYYDVRSLYPFIMKSSPMEICAA
gOryzae3 : AFIR...RGYYGGHA...DTYKP.....YGDNLYYYDVNSLYPFAIENISQAPIHL
gMarchantia1 : QFLR...KAFYGGHT...DVYKP.....YGENLYVYDSNSLYPFSMT.KK.....
pCLK1 641: NDIH...EAYYGGRV...EVFNPIIMADSTKSYYYYDVNSLYPFASIN.DIPGLKC
pBgh 606: DDIH...SAYYGGRV...EVYNPVIAPGEVK..YYYDGNSLYPYASLN.EIPGVKY
phi29 219: KEVR...YAYRGGFTWLNDRFKEK...EIGEG.MVFDVNSLYPAQMYSRLLPYGEP


TPR1
gLpDpofull 600: DGINKPFLPSKHPVED.....TLIFPEG.TFFGTYFT.EDLKYAKSIGYEVIIT
FNzDpo : DGINKPFLPSKHPVED.....TLIFPEG.PFFGTYFT.EELKYAKSIGYEVIIT
FIp4Dpo : DGINKPFLPSKHPVED.....TLIFPEG.PFFGTYFT.EELKYAKSIGYEVIIT
gDaucusfull 329: DSIKKPFLPYREKVNG.....SLIFPTG.YFVGVYYS.EELKYARDLGYTIYPL
pBrassica 604: THIHKPFLPY.KEHD.....ILRFPTG.IWVGVYFS.EELKLAKSLGYTIIPI
pBetav 554: KSIKKPFLPYFNKKDG.....TLIFPTG.TFVGVYYS.EELKYARDIGYTIIPI
gBetavfull 384: ESIVRPFLPYHKVKDP.....TLIFPTG.EWIGVYYT.EELKYAKSLGYTVIPI
gBetav2: ETIKRPFLPFRDKKRG.....LIFPTG.SFFGVYYS.EELKYARDIGYTVIPL
gBetav3 : ETRKRPFLPFRDKKRG.....LIFPTG.SFFGVYYS.EELKYARDIGYTVIPL
pS1 524: KHIKKPLLPYKKDDG.....TIIFPTG.RFLGVYFS.EELKYAVSLGYKVIPI
gZmNB1full 349: KHIKKPLLPYKKDDG.....TIIFPTG.RFLGVYFS.EELKYAVSLGYKVIPI
gZmNB2full 535: ADMHKPFLPYKQSDG.....TLIFPTG.RFIGVYFS.EELKYAVSLGYKVIPI
gZmN1 : .....WY...DN.....LPHG.RLLGVYFS.EELKYAVSLGYKVIPI
gOryzae2 : ETMNKPFLPY.....FIGVYYS.EKLQYADIFGYKVRPL
gMarchantia1 : LDFNRPFLPFRLEQLK.....YLIFPTG.NWKGTYYS.KKLKYAKSLG.....
pCLK1 706: DNLYLGLLPKRTETS.....LIFPGG.AWEGWYFS.EELKFAVAHGYEVEII
pBgh 678: SGNYLGLLPTRTKTN.....LIFPVG.EWFGWYFS.EELKFAKDNGYVIEVI
phi29 290: FELKEGYIPTIQIKRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEH.YDLYNV


Pol IIa
Pol IIb
gLpDpofull 668:
FNzDpo :
FIp4Dpo :
gDaucusfull 397:
pBrassica 671:
pBetav 631:
gBetavfull 454:
gBetav2 :
gBetav3 :
gBetav4 :
gBetav5
gBetav6
gBetav7
gBetav8
pS1 600:
gZmNB1full 416:
gZmNB2full 601:
gZmN1 :
gSecale
pCLK1 781:
pBgh 740:
phi29 368:
754:
:
:
481:
757:
716:
546:
:
:
:
:
:
:
700:
498:
680:
:
:
883:
847:
425:

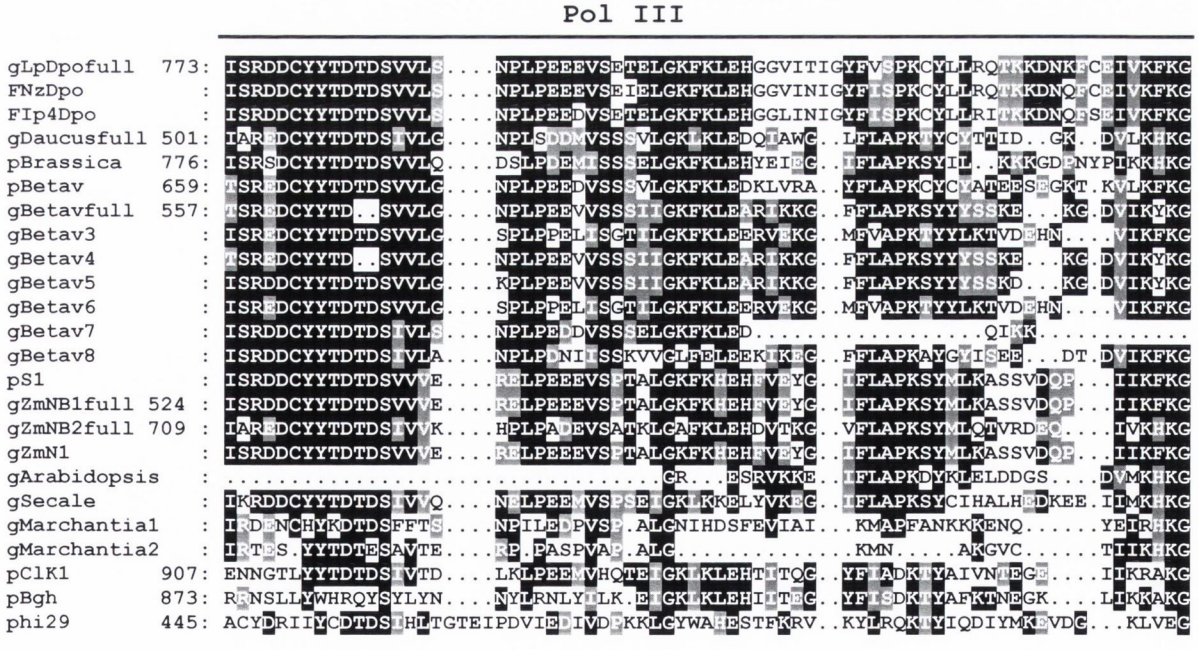
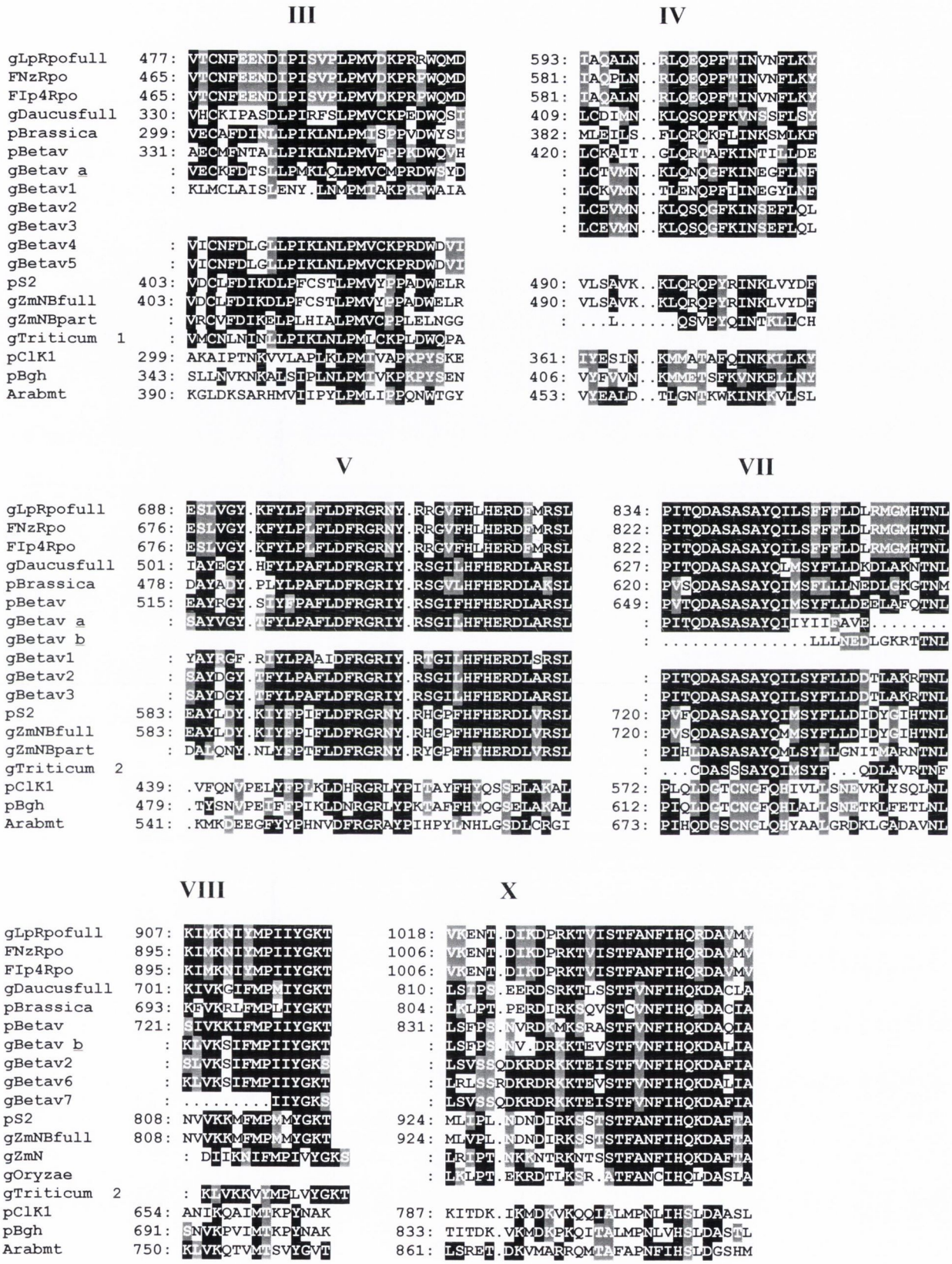

Pol III
gLpDpofull 773:
FNzDpo :
FIp4Dpo :
gDaucusfull 501:
pBrassica 776:
pBetav 659:
gBetavfull 557:
gBetav3 :
gBetav4 :
gBetav5 :
gBetav6 :
gBetav7 :
gBetav8 :
pS1 :
gZmNB1full 524 :
gZmNB2full 709 :
gZmN1 :
gArabidopsis :
gSecale :
gMarchantia1 :
gMarchantia2 :
pClK1 907:
pBgh 873:
phi29 445:

Figure 4.28 Comparison of the putative DNA polymerase ORF1 in the mitochondrial genome of *L.perenne* with other plant

Comparison of sterile line of *L. perenne* CMS9B290–(gLpDpofull) with the most likely conserved domains of linear plasmid like sequence in two fertile lines of *L. perenne*; Irish pasture *FIp4*–FIp4Dpo, New Zealand *FNz* –FNzDpo. and from the mitochondria of other plant and fungi. DNAPol encoded reading frames in the mitochondrial genome of *Daucus carota* AY521591 [gDaucus(full)], *Beta vulgaris* Z34298 [gBetav(full)], *Beta vulgaris* BA000009 [gBetav 1- *orf409*, gBetav 2- *orf*202, gBetav 4- *orf*310, gBetav 5- *orf*192, gBetav 6- *orf*290, gBetav 7- *orf* 135], *Beta vulgaris* BA000024 [gBetav 3- *orf*496, gBetav 8- *orf* 284], *Zea mays* (NB) AY506529 [gZmNB1(full)- *orf*734, gZmNB2(full)- *orf*911, gZmNB3 (158600-157800)], *Zea mays* (N) X07041 [gZmN1(2,858-4,749)], *Oryza sativa* AB076665 [gOryza 1 (453700-453930), gOryza 2 (259800-259200), AB076666; gOryza 3 (112150-114000)], *Marchantia polymorpha* M68929 [gMarchantia 1 (89200-92700), gMarchantia 2 (47700-48300)], *Arabidopsis thaliana* Y08502 [gArabidopsis (115500-115900)], *Secale cereale* (X74133) [gSecale], *Nicotiana tobaccum* BAD83421 [gNtobaccum- *orf297*]. Similar DNAPol amino acid sequence encoded by free living linear mitochondrial plasmids in the three plant sp. *Brassica napus* AB073400 [pBrassica], *Zea mays* (S) X02451 [pS1], *Beta vulgaris* Y10854 [pBetav], two fungal linear plasmids *Claviceps purpurea* P22373 [pClK1], *Blumeria graminis* AY189817 [pBgh], and the bacteriophage phi29 of *Bacillus subtillus* P03680 [phi29]. Numbers of amino acids are indicated only for complete DNA Pol reading frames containing all of the conserved domains. Amino acids with an identity of three or more are shaded. The dark shade quantifies the residue with the highest identity, and residues with the second most identity are given the lighter shade. Note: The word full highlights DNA polymerase *orfs* containing all the conserved domains integrated within the mitochondrial genome.


III
IV
V
VII
VIII
X
gLpRpofull
FNzRpo
FIp4Rpo
gDaucusfull
pBrassica
pBetav
gBetav a
gBetav1
gBetav2
gBetav3
gBetav4
gBetav5
pS2
gZmNBfull
gZmNBpart
gTriticum 1
pClK1
pBgh
Arabmt
gBetav b
gBetav6
gBetav7
gZmN
gOryzae
gTriticum 2

XI

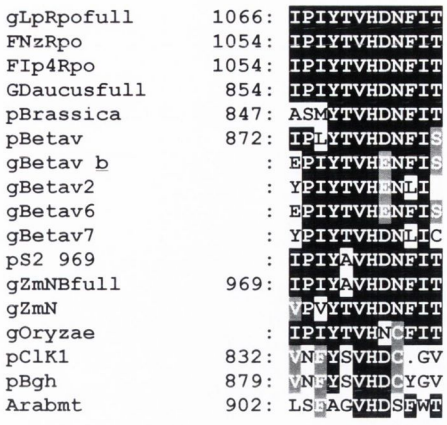


Figure 4.29 Comparison of the putative RNA polymerase ORF2 in the mitochondrial genome of *L. perenne* with other plant systems

Comparison of a sterile line of *L. perenne* CMS9B290 (gLpRpofull) with the most likely conserved domains of plasmid like sequence in two fertile lines of *L. perenne*; Irish pasture *FIp4* (FIp4Rpo), New Zealand *FNz* (FNzRpo) and from the mitochondria of other plant and fungi. RNAPol encoded reading frames in the mitochondrial genome of *Daucus carota* AY521591 [gDaucus(full)], *Beta vulgaris* Z34298 [gBetav a (463-2152), gBetav b (2,139-3,235)], *Beta vulgaris* BA000009 [gBetav 4- *orf187,* gBetav 6- *orf348*], *Beta vulgaris* BA000024 [gBetav 1- *orf*329, gBetav 2- *orf598*, gBetav 3- *orf*352, gBetav 5- *orf185,* gBetav 7- *orf217*], *Zea mays* (NB) AY506529 [gZmNB(full)- *orf*1159, gZmNB (363000-360000)], *Oryza sativa* AB076665 [gOryza (208900-211500)], *Triticum aestivum* AP000829 [gTriticum1- orf349], Z11889 [gTriticum2 (850-1150)]. Similar RNAPol amino acid sequence encoded by free living linear mitochondrial plasmids in the three plant sp. *Brassica napus* AB073400 [pBrassica], *Zea mays* (S) J01426 [pS2], *Beta vulgaris* Y10854 [pBetav], two fungal linear plasmids *Claviceps purpurea* P22373 [pClK1], *Blumeria graminis* AY189817 [pBgh], and the nuclear encoded single subunit mtRNA polymerase in *Arabidopsis thaliana* AAG51592 [Arabmt]. Numbers of amino acids are indicated only for complete RNA Pol reading frames. containing all of the conserved domains. Amino acids with an identity of three or more are shaded. The dark shade quantifies the residue with the highest identity, and residues with the second most identity are given the lighter shade. Note: The word full highlights RNA polymerase *orfs* containing all of the conserved domains integrated within the mitochondrial genome.

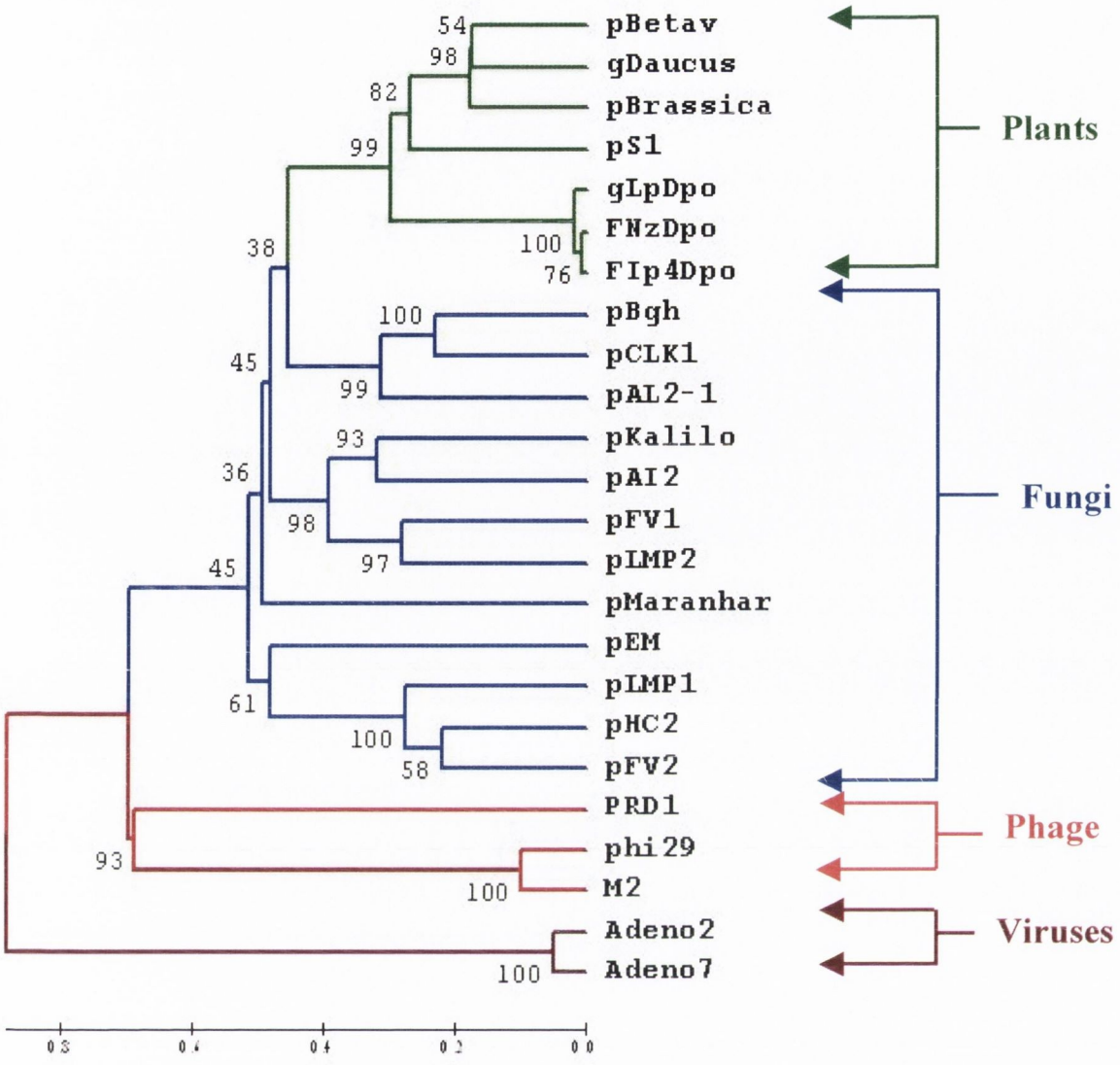


Figure 4.30 Relationship of Dpo in *L. perenne* with DNA B Polymerases

Phylogenetic relationship of the putative ORF1 DNA Polymerase amino acid sequences in CMS *L. perenne* with the conserved DNA Polymerase of linear plasmids in plants, fungi, and DNA Polymerase in phages and viruses (Kempken *et al* ., 1992). Multiple alignment with ClustalW (Thompson *et al*., 1994) was used to generate a phylogenetic tree; *L. perenne* gLpRpo (CMSi), FNzRpo, FIp4Rpo, S1 plasmid in *Zea mays* (X02451), 11.6 kb plasmid in *Brassica napus* (AB073400), 10.4 kb plasmid in *Beta vulgaris var.maritima* (Y10854), *Daucus carota (AAS15054),* Kalilio plasmid in *Neurospora crassa* (X52106), Maranhar plasmid in *Neurospora intermedia* (X55361), pHC2 plasmid in *Hebeloma circinans* (CAA72280), pEM plasmid in *Agaricus bitorquis* (CAA44800), pCLK1 plasmid in *Claviceps purpurea* (P22373), pLMP1/2 plasmid in *Pleurotus ostreatus* (AF126285), pFV1/2 plasmid in *Flammulina velutipes* (AB028633, AB028634), pAI2 plasmid in *Ascobolus immerus* (S05342), pAL2-1 in plasmid *Podospora anserina* (QQ1529), bacteriophage PRD1 DNA Polymerase (M69077), phi29 bacteriophage DNA Polymerase in *Bacillus subtilus* (P03680), Adenoviruse Adeno 2, 7 DNA Polymerase (AAA92206, AA26770), M2 phage (P19894). The length of the branches estimates the distances and the scale bar represents the number of amino acid substitutions per site.

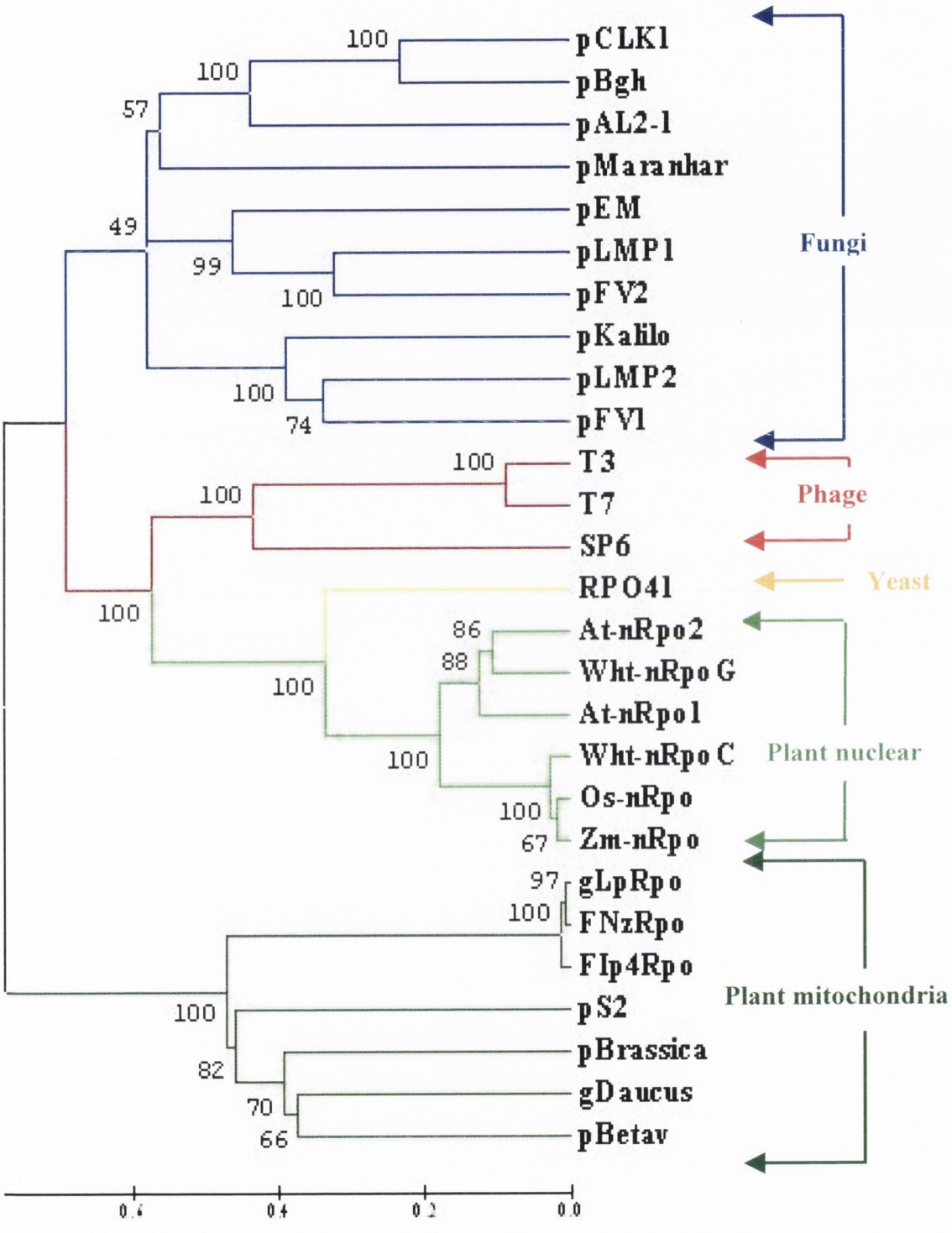

pCLK1
pBgh
pAL2-1
pMaranhar
pEM
pLMP1
pFV2
pKalilo
pLMP2
pFV1
T3
T7
SP6
RPO41
At-nRpo2
Wht-nRpo G
At-nRpo1
Wht-nRpo C
Os-nRpo
Zm-nRpo
gLpRpo
FNzRpo
FIp4Rpo
pS2
pBrassica
gDaucus
pBetav
100
100
57
49
99
100
100
74
100
100
100
86
88
100
100
100
67
97
100
100
82
70
66
0.6
0.4
0.2
0.0
Fungi
Phage
Yeast
Plant nuclear
Plant mitochondria

Figure 4.31 Relationship of Rpo in *L. perenne* with single subunit RNA Polymerases

Phylogenetic relationship of ORF2 RNA Polymerase amino acid sequences in CMS *L. perenne* with the conserved single subunit RNA Polymerase of linear plasmids in plants (mitochondrial in origin), fungi, and single subunit RNA Polymerase in plants (nuclear in origin), phages and yeast (Kempken *et al* ., 1992). Multiple alignment of RNA Polymerases by ClustalW (Thompson *et al*., 1994) used to generate a phylogenetic tree; *L. perenne* gLpRpo (CMSi), FNzRpo, FIp4Rpo, pCLK1 *Claviceps purpurea* (P22372), Kalilo *Neurospora crassa* (X52106), Maranhar *Neurospora intermedia* (X55361), pEM *Agaricus bisporus* (AAD11785), pAL2-1, *Podospora anserinia* (S26945), pLMP1/2 *Pleurotus ostreatus* (AF126285), pFV1/2 in *Flammulina velitupes*(AB028633, AB028634) linear plasmids in plant were; S2 - *Zea mays* (J01426), 11.64kb plasmid *Brassica napus* (AB073400), 10.4 plasmid, *Beta vulgaris var. maritima* (AB073400), *Daucus carota* (AAS15052) mitochondrial. RNA. Polymerase in yeast; RPO41 *Saccharomyces cerevisiae* (.NP116617), bacteriophage SP6 RNA Polymerase (CAA68288), T3 and T7 bacteriphage (RNBPT3, M38308), nuclear encoded ssRNA Polymerases in plant Wht-nRpo - *Triticum* (C, F091839; G, F091838), At-nRpo, - *Arabidopsis* (1, Y08137; 2, AJ001037), Zm-nRpo - *Zea mays* (AF127021), Os-nRpo -*Oryzae sativa* (BAC76605). The length of the branches estimates the distances and the scale bar represents the number of amino acid substitutions per site.

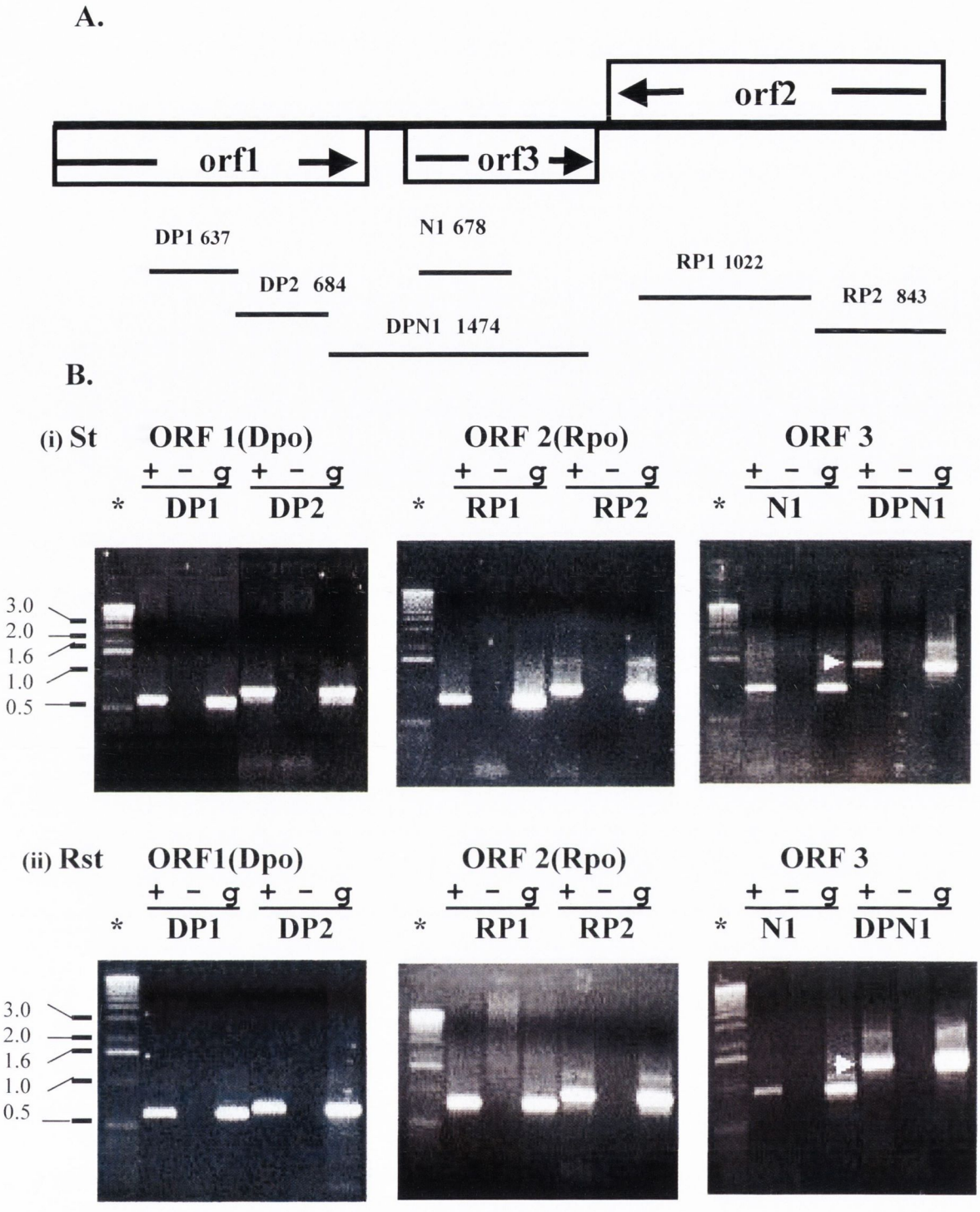


Figure 4.32 Evidence of Transcripts for ORF1, ORF2, ORF3 by RT-PCR

A. Location of the RT-PCR products spanning each of the three orf's 1, 2, 3 in the 9.6 kb CMSi mitochondrial region. **B.** Indicates RT-PCR amplified products for orf 1, 2, 3 in **(i)** a sterile (St) and **(ii)** a restored (Rst1) *L. perenne* line. The amplified DP1, DP2, DPN1, N1, RP1, and RP2 PCR products are acheived with the use of reverse transcriptase **+**, and controls used are no reverse transcriptase **-** (negative control) and genomic DNA (positive control). The clear arrow highlights the 1.474 kb orf1(Dpo)-orf3 (DPN1) co-transcript. The *asterick* symbol * is the 1 kb size marker .

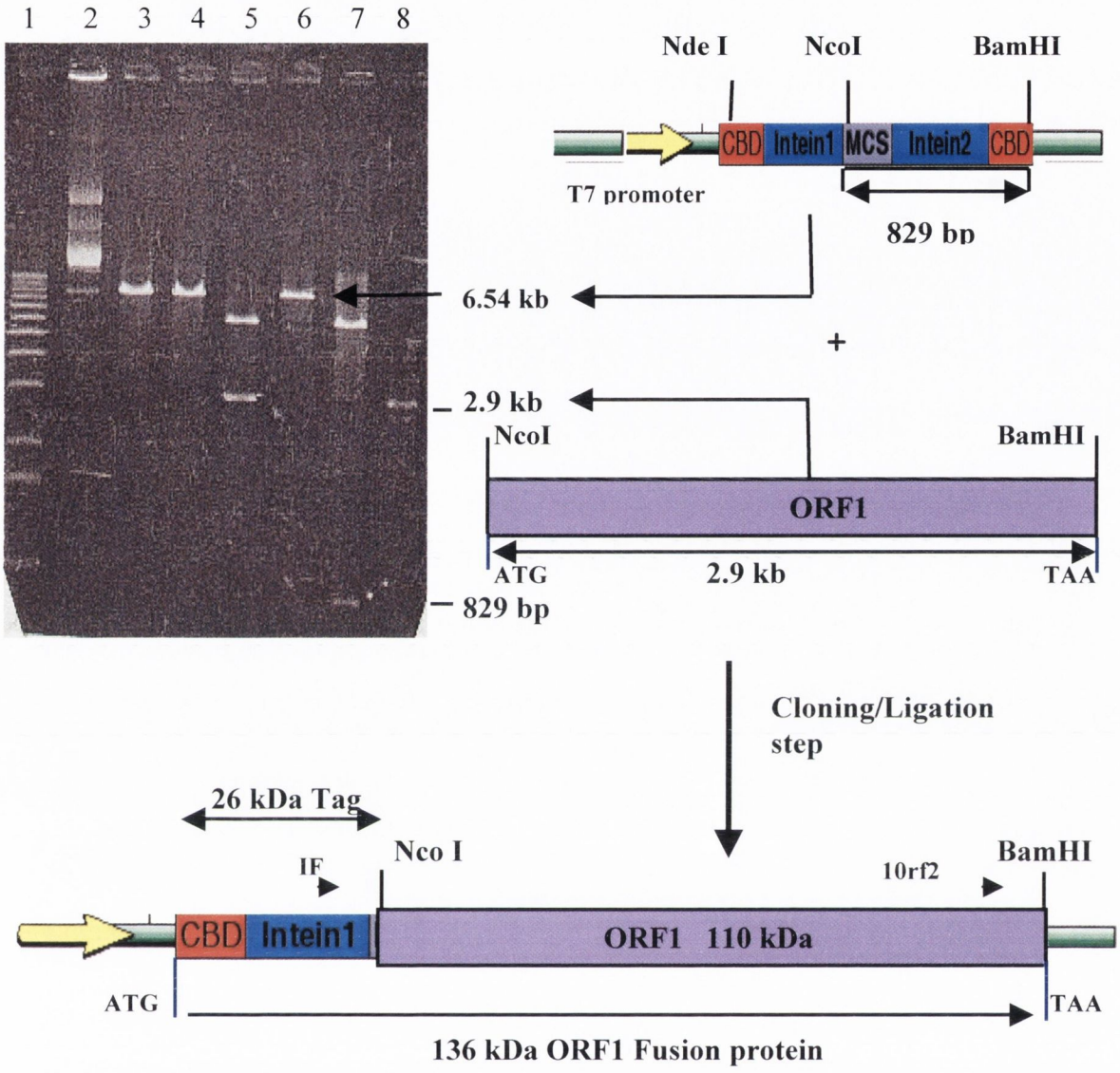


Figure 4.33 Confirmation of the pTwinILpDpo(orf1) construct by restriction digest.

Lane 1 is the 1 kb ladder, lane 2 is uncut pTwinIorf1-2.9 construct, lane 3 *Bam*HI digested pTwinIorf1-2.9, lane 4 *Nco*I digest pTwinI orf1-2.9, lane 5 is a double *Bam*HI */Nco*I digest pTwinorf1-2.9, lane 6 *Nde*I digest pTwinILpDpol, lane 7 is a double *Bam*HI/*Nco*I digest of pTwinI impact vector with two band sizes of 6.54 kb and 829 bp respectively, lane 8 is the orf1 2.9 kb fragment. The primers IF and 10rf2 highlighted by arrow are used to confirm sequence at either junction.

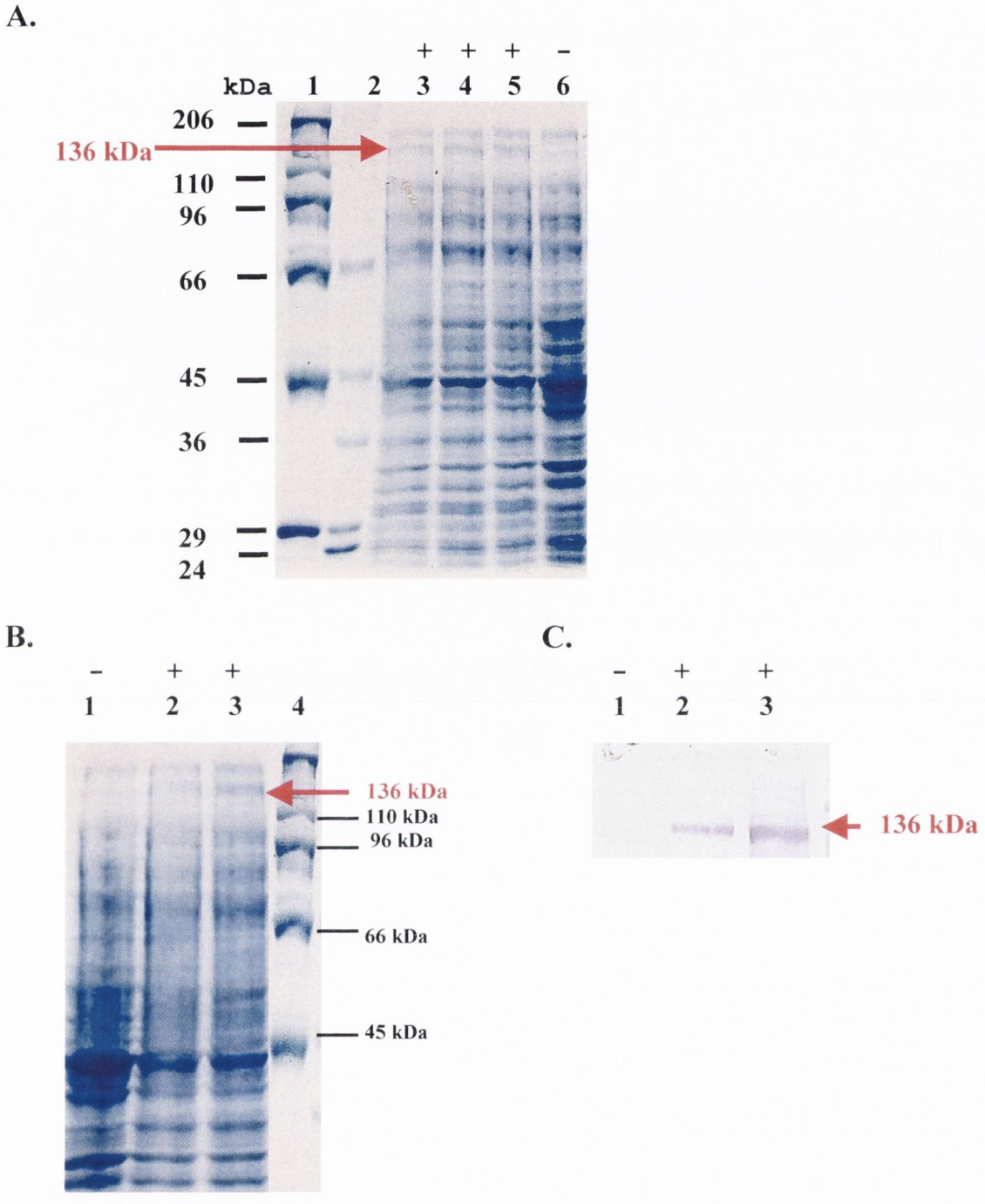


Figure 4.34 Expression of the 136 kDa fusion protein in pTwinILpDpo(ORF1) construct

A Lanes 1 and 2 are protein markers of high and low range respectively, lanes 3, 4, 5 is the induced (+) whole cell protein containing the expressed 136 kDa fusion protein, lane 6 is the uninduced (−) whole cell control protein sample. **B** Lanes 1 is the uninduced (−) soluble protein fraction, lanes 2 and 3 is the induced (+) soluble protein fraction containing the 136 kDa fusion protein lane 4 is the high range molecular marker.

C Western blot of the SDS-PAGE in Figure 4.34B which has been probed with the commercially supplied (NEB) pTwin impact Chitin binding antibody (Ab).

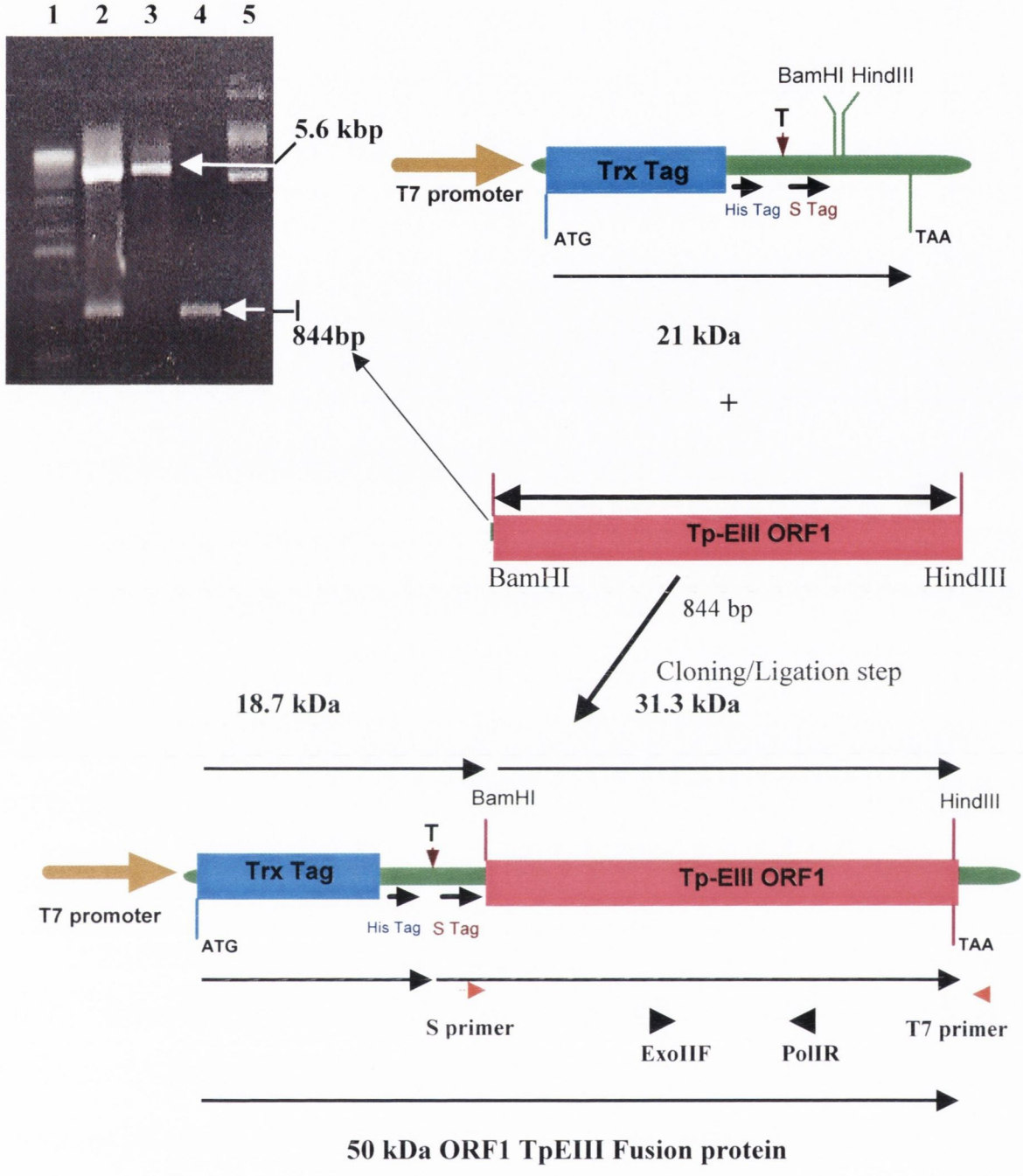


Figure 4.35 Illustrates the insertion of the LpDpoTpEIII 844bp size fragment into the pET32a expression vector.

Lane 1 is the 1.0 kb DNA ladder, lane 2 is a *Bam*HI/*Hind*III digest of the pet32a TpEIII clone, lane 3 is a *Bam*HI/*Hind*III digest of the pet32a vector, lane 4 is the amplified TpEIII 844 bp size fragment. Lane 5 is the undigested pet32aTpEIII clone.

Promoter used is the T7, the S primer, T7 primer, ExoIIF and Pol IR primers are used to confirm sequence of TpEIII insert, the Tags highlighted are the His, S and Trx tags.

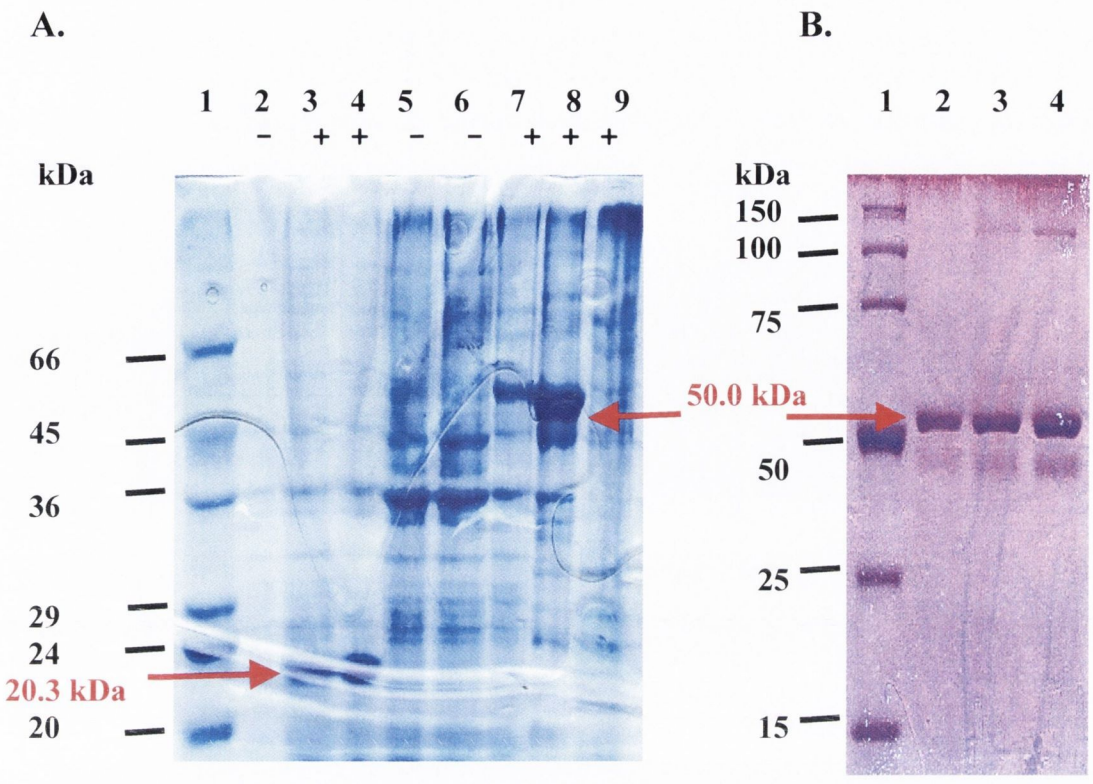


Figure 4.36 Protein Expression of pET32LpDpoTpEIII

A. Illustrates the expression of the pET32LpDpoTpEIII 50 kDa fusion protein under an incubation temperature of 30°C and an induction concentration of 1mM IPTG on a 10% P.A.G.E. Lane 1 is the low range molecular protein standard, lane 2 is uninduced whole cell protein pET32LpDpoTpEIII at T=0, lanes 3 to 4 are the induced whole cell protein of pET32a vector at T= 2hr. and 4hr. respectively. Lanes 5 to 6 is uninduced whole cell protein of pET32a vector at T=4hr. and pET32LpDpoTpEIII at T=4hr., lane 7 is induced whole cell protein of pET32TpEIII at T=4hr., lane 8 is induced insoluble fraction of pET32LpDpoTpEIII at T= 4hr., lane 9 is the induced soluble fraction of pET32LpDpoTpEIII at T=4hr. **B.** Illustrates the purified insoluble 50.0 kDa fusion protein that was used to generate polyclonal antibodies in rabbit. Lane 1 is the novagen molecular protein standard , lane 2 is a portion (1.5 µg) of protein used in the 2nd booster injection for the rabbit namely – 100 µg, lane 3 is a portion (1.4 µg) of protein used in the 1st booster injection for the rabbit – 150 µg, and lane 4 is a portion (2µg) of protein used in the initial injection of the rabbit – 200 µg.

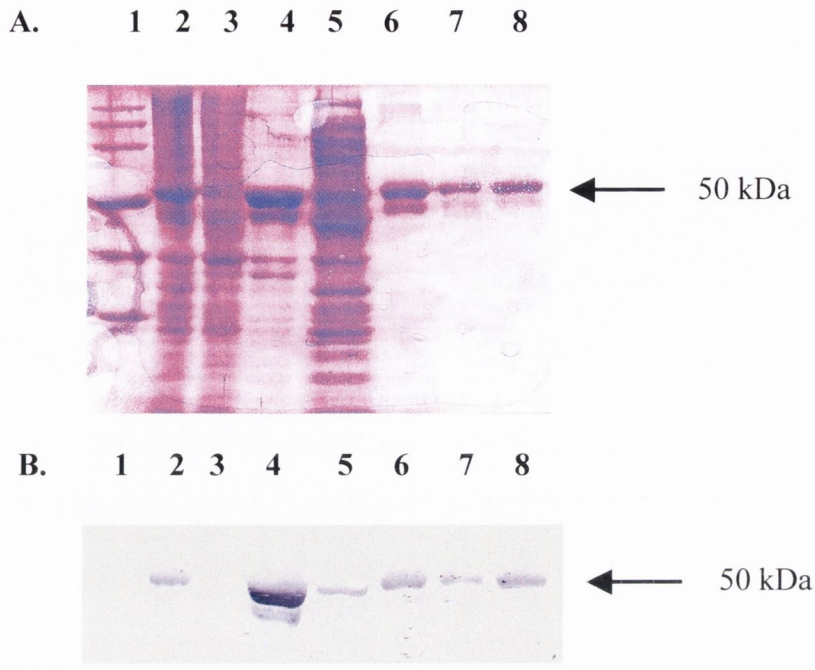


Figure 4.37 Western blot analysis of the pET32a recombinant expression system probed with the anti-Trx-LpDpoTpEIII antibody

A. Is total leaf protein run on a 10% SDS-PAG stained in coomassie blue. Lane 1 is the protein marker, lane 2 is induced whole cell protein, lane 3 is uninduced whole cell, lane 4 is the induced insoluble fraction of pET32aLpDpo, lane 5 is the soluble fraction, lane 6 to lane 8 are purified protein fractions of the 50.0 kDa fusion protein. **B.** Is an immunoblot of 10% SDS-PAG above in **A,** and is probed with the Trx-LpDpoTpEIII antibody.

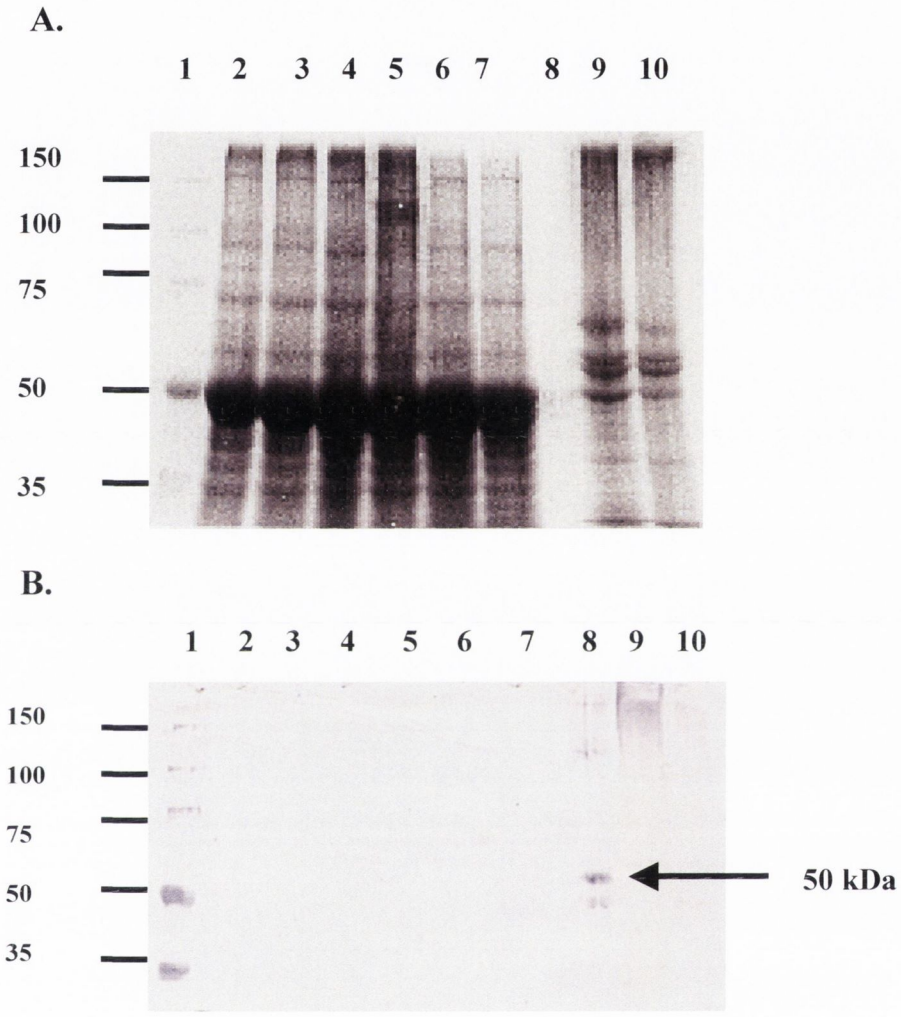


Figure 4.38 Western blot of a 10% SDS-PAGE of leaf protein probed with anti-Trx- LpDpoTpEIII antibodies

A. Lane 1 is the protein marker, lane 2 and 3 is fertile wild type protein (80 µg), lane 4 is restored protein (100 µg), lane 5 to 7 is sterile protein (100 µg), lane 8 is Insoluble protein (40 ng), lane 9 is sterile mitochondrial protein (50 µg), lane 10 is wild type mitochondrial protein (50 µg). **B.** Is a Western blot of 10% Coomassie gel in A. probed with aniti Trx-LpDpoTpEIII polyclonal anti-body.

Table 4.1 Synonymous and non-synonymous amino acid substitutions for Dpo in *L. perenne* sterile line CMS9B290, and in the fertile lines FNz and FIp4

Codon No	Lp-CMSi	FNz	FIp4	Domain
13	GCC (A)		GCT	
19	GAT (D)		GAA (E)	
25	GAA (E)		GAT (D)	
48	ATC (I)		ACC (T)	
51	TCC (S)	TAC (Y)	TAC (Y)	
66	GAC (D)		GAT	
71	CCA (P)		CCT	
78	ACA (T)	ACT	ACT	
79	TCA (S)	TCT	TCT	
107	CTG (L)	ATG (M)	ATG (M)	
116	ATG (M)	ATA (I)	ATA (I)	
120	CAG (Q)	CCG (P)		ExoI
123	GCA (A)		GAG (E)	ExoI
125	CCT (P)	CCG	CAG (Q)	ExoI
128	CAG (Q)	CTG (L)		
196	CTG (L)	ATG (M)	ATG (M)	
210	CTG (L)		CAG (Q)	
216	GTC (V)		GTG	
218	TTT (F)		TTG (L)	ExoII
238	AAT (N)		ATT (I)	
241	TAC (Y)		TAT	
339	CTA (L)	CTG		
340	GGT (G)	GGA		
341	GTC (V)		GTA	
358	ATC (I)	ATA		
375	GAA (E)	CAA (Q)	CAA (Q)	
432	TAC (Y)		TTC (F)	Pol I
444	ACT (T)		AGT (S)	
485	ACT (T)	CCT (P)	CCT (P)	TPR1
494	GAC (D)	GAA (E)	GAA (E)	TPR1
502	GGG (G)	GGC	GGC	TPR1
529	AAA (K)		GCA (A)	Pol IIa
536	GCT (A)		GCA	Pol IIa
541	GTA (V)	ATT (I)	GTT	Pol IIa
594	ATA (I)	GTA (V)	GTA (V)	
615	ACA (T)	ACT		
629	ATC (I)	ATG (M)	ATG (M)	Pol IIb
658	GAG (E)		GAT (D)	Pol III
662	ACA (T)	ATA (I)		Pol III
665	GGC (G)	GGA		Pol III
672	GGA (G)		GGT	Pol III
674	GTA (V)		TTA (L)	Pol III
676	ACC (T)	AAC (N)	AAC (N)	Pol III
681	GTA (V)	ATA (I)	ATA (I)	Pol III
690	CAA (Q)		ATC (I)	Pol III
692	AAG (K)		AAA	Pol III
696	AAG (K)	CAG (Q)	CAG (Q)	Pol III
698	TGT (C)		AGT (S)	Pol III

717	GAC (D)		GAG (E)	Pol III
720	ATC (I)		ACA (T) d	Pol III
724	GAG (E)		CAG (Q)	
726	ATA (I)	ACA (T)	ACA (T)	
741	TTC (F)		TTT	
744	GTT (V)	ATT (I)	ATT (I)	
758	AAA (K)		AAT (N)	
761	AAA (K)		GTA (V)	
762	GCA (A)		CCA (P)	
772	GAT (D)	GGT (G)		
781	AAT (N)		AAG (K)	
784	AGA (R)		AGT (S)	
795	GTC (V)		GTG	
798	CTG (L)		ATG (M)	
799	TCA (S)		CAA (Q)	
801	CTA (L)		ATC (I)	
802	AGT (S)		AAA (K)	
803	AGG (R)		AAG (K)	
822	GAT (D)		TCA (S)	
827	GAC (D)	AAC (N)		

LpCMSi: is the sterile triplet codon, and name of codon is enclosed in brackets.

FNz: is the fertile triplet codon, Non-synonomous alteration in amino acid relative to LpCMSi are highlighted by codon name in bracketts. Synonomous triplet codons are highlighted by the absenence of brackets. Synonomous substitutions are highlighted in red.

FIp4: is the fertile triplet codon. Non-synonomous alterationsin amino acid relative to LpCMSi are highlighted by the codon name in brackets. Synonomous triplet codons are highlighted by the absenence of brackets. Synonomous substitutions are highlighted in red.

Table 4.2 Synonymous and non-synonymous amino acid substitutions for Rpo in the *L. perenne* sterile line CMS9B290, and in the fertile lines FNz and FIp4

Codon No.	Lp-CMSi	FNz	FIp4	Domain
3	ATT (I)	AAA (K)	AAA (K)	
4	AAA (K)	CAA (Q)	CAA (Q)	
11	GCG (A)	GCA	GCA	
12	ACA (T)	ATA (I)	ATA (I)	
14	TCT (S)	GAT (D)	GAT (D)	
21	AAT (N)	ACT (T)	ACT (T)	
26	CCC (P)	CCA	CCA	
33	ATG (M)	AAC (N)	AAC (N)	
34	GAA (E)	AAT (N)	AAT (N)	
40	ATT (I)	AAT (N)	AAT (N)	
47	AAG (K)	ATG (M)	ATG (M)	
60	ACA (T)	ACT	ACT	
61	AAA (K)	TTA (L)	TTA (L)	
63	GCT (A)	ACT (T)	ACT (T)	
65	CAG (Q)	AAG (K)	AAG (K)	
76	ACA (T)	ATA (I)	ATA (I)	
97	TAT (Y)	TTT (F)	TTT (F)	
98	CTA (L)	CTT	CTT	
101	GAT (D)	AAT (N)	AAT (N)	
107	CAA (Q)	CAT (H)	CAT (H)	
113	GAA (E)	GCA (A)	GCA (A)	
145	GAA (E)	GCA (A)	GCA (A)	
149	GGA (G)	TCC (S)	TCC (S)	
153	ACT (T)	ACA	ACA	
157	TAT (Y)	TTC (F)	TTC (F)	
178	CAA (Q)		AAA (K)	
186	GCT (A)		GCA	
194	ACT (T)	ACG	ACG	
200	AAT (N)		AGT (S)	
248	TTG (L)		ATG (M)	
254	GGA (G)	GGT	GGT	
265	AAA (K)	CAA (Q)	CAA (Q)	
272	ATA (I)		ATG (M)	
280	GAT (D)		GAG (E)	
284	AAA (K)		CAA (Q)	
286	TTA (L)	ATA (I)		
306	ATT (I)		ATC	
307	TGT (C)		GCG (A)	
312	AGA (R)		AGG	
363	TAT (Y)		AAT (N)	
371	AAA (K)	CAA (Q)		
373	GAA (E)		AAA (K)	
381	TCT (S)		GAT (D)	
383	CAA (Q)		CTA (L)	
384	AAT (N)		AAA (K)	
390	AAA (K)		AAT (N)	
393	TCT (S)		GAT (D)	
399	TCT (S)	GCT (A)	GAT (D)	
407	CTA (L)	ATA (I)		

417	AGA (R)		ATA (I)	
420	AAA (K)	CAA (Q)		
439	AAG (K)		AGG (R)	II
442	CCC (P)		CAC (H)	
454	ACA (T)		ACT	
459	TAC (Y)		TTC (F)	
461	AGT (S)		ACT (T)	
464	TCC (S)		GCC (A)	
489	CGA (R)	CCA (P)	CCA (P)	III
495	TAC (Y)		TTC (F)	
496	CGT (R)	CAT (H)	CAT (H)	
498	GAA (E)		CAA (Q)	
501	TAC (Y)	TTT (F)	TTT (F)	
504	AAT (N)		TAT (Y)	
510	AAG (K)		CAA (Q)	
516	CTC (L)	TCC (S)	TCC (S)	
522	ACA (T)		ACT	
525	GCC (A)		ACC (T)	
526	CAG (Q)		CTG (L)	
527	CCA (P)	CTA (L)	CTA (L)	
534	ATT (I)		ATG (M)	
574	TGC (C)		AGC (S)	
575	AAT (N)	TAT (Y)	TTT (F)	
577	GAT (D)		GAA (E)	
585	GCC (A)	CCC (P)		IV
609	AGT (S)	GGT (G)	GGT (G)	
630	GTA (V)		ATA (I)	
644	AAA (K)	ACA (T)		
653	TAT (Y)	AAT (N)	AAT (N)	
720	AAT (N)		GAT (D)	
725	AAA (K)		ATA (I)	
735	GAG (E)		GAC (D)	
753	TTC (F)	CTC (L)	CAC (H)	
757	ATG (M)	AAG (K)	AAG (K)	
760	CCC (P)	CCG	CCG	
771	GAA (E)		GTT (V)	
772	AAG (K)		CAG (Q)	
783	GAA (E)	AAA (K)		
788	GAT (D)		AAT (N)	
809	CAT (H)		CAA (Q)	VI
871	TGT (C)		TAT (Y)	
873	ACA (T)		TCA (S)	
878	ATC (I)		AAC (N)	
880	GAA (E)	GCA (A)		
884	ATG (M)		AAG (K)	
890	ATC (I)	ATA	ATA	
912	AAT (N)	AAG (K)	AAG (K)	VIII
925	AAT (N)		GAT (D)	
928	TAT (Y)		ATT (I)	
947	AAA (K)		AAT (N)	
964	ACG (T)	ACT	ACT	
976	TTG (L)		TGG (W)	IX

1000	AGA (R)		AAA (K)	
1049	CAG (Q)	AAG (K)	AAG (K)	
1072	TAT (Y)	TCT (S)	TCA (S)	
1082	TTG (L)	TTC (F)	TTC (F)	
1104	GCA (A)		GCC	
1133	TTA (L)		GTA (V)	
1144	CAC (H)		AAC (N)	
1148	TCT (S)		TTT (F)	
1152	CAA (Q)		CCA (P)	
1157	GAA (E)		GAT (D)	
1159	GTT (V)		TTT (F)	
1161	AGA (R)		GGA (G)	
1163	GTA (V)	GCA (A)	GCA (A)	
1172	CAA (Q)		CTA (L)	
1179	AAT (N)		GAT (D)	
1180	GAC (D)		AAC (N)	
1187	GAC (D)		TCC (S)	
1188	TTC (F)		ATC (I)	
1192	GTG (V)	GAG (E)	GAG (E)	
1193	GTC (V)		GGC (G)	
1212	TTA (L)		AAA (K)	
1213	AAA (K)	AAT (N)	AAT (N)	
1216	AGC (S)		CGC (R)	
1217	ACC (T)	TCC (S)	TCC (S)	
1240	GAA (E)	AAA (K)	AAA (K)	
1249	GAA (E)	CAA (Q)	CAA (Q)	
1254	TTC (F)		GTC (V)	
1266	TAT (Y)		TTT (F)	

LpCMSi: is the sterile triplet codon, and name of codon is enclosed in brackets.

FNz: is the fertile triplet codon, Non-synonomous alteration in amino acid relative to LpCMSi are highlighted by codon name in bracketts. Synonomous triplet codons are highlighted by the absenence of brackets. Synonomous substitutions are highlighted in red.

FIp4: is the fertile triplet codon. Non-synonomous alterationsin amino acid relative to LpCMSi are highlighted by the codon name in brackets. Synonomous triplet codons are highlighted by the absenence of brackets. Synonomous substitutions are highlighted in red.

CHAPTER 5

Expression of the CMS-associated *orfC9* gene in Tobacco

5.1 Introduction

F1 hybrid plants perform better than inbred lines in respect of vigour, disease resistance and yield. To produce F1 hybrids on a commercial scale it is necessary to devise an efficient mechanism to minimize self-fertilization. As described in Chapter 3, in the case of CMS, male fertility is suppressed by naturally-occurring mitochondrial mutations and fertility can be restored by nuclear *Rf* genes. With the advent of transgenic technologies, an alternative approach became possible i.e. to engineer male sterility using transgenes. The strategies explored, so far, fall into two categories: those based on transgenes of non-plant origin, the products of which interfere with some stage of pollen development; those based on transgenes derived from well-characterized CMS-determining genes. However, the ability to produce a male sterile phenotype must also be complemented with a strategy to restore male fertility, since hybrid seeds are typically required to produce fertile plants. Transgenic restoration of fertility has proved to be somewhat more problematic.

5.1.1 Engineered male sterility based on transgenes of non-plant origin

The first reported transgenic approach for engineering male sterility using a non-plant transgene was based on the use of an anther-specific promoter, Ta29, to drive expression of the *Bacillus amyloliquefaciens* gene coding for RNAase T1 (also known as barnase) in tobacco plants (Mariani *et al.*, 1990). Barnase was specifically expressed in the anther tapetal cell layer, resulting in the ablation of pollen production and in consequence, male sterility. Restoration of fertility was achieved in this system by regulated expression of a second transgene encoding the RNAase TI inhibitor, barstar (Mariani *et al.*, 1990). More recently, transformation of *Nicotiana*, *Arabidopsis*, *Brassica* and *Solanum* plants with a barnase transgene under control of the pea anther primordium-specific promoter

PsEND1 produced a strong male-sterile phenotype by interfering with the earliest stages of anther development (Roque *et al.*, 2007). Most recently rearrangements in the plant mitochondrial genome was achieved by suppression of *Msh1*, by transformation with an RNAi transgene, thus resulting in induction of partial male sterility in transgenic plants of tobacco (Shandu *et al.*, 2007).

Engineered male sterility has also been achieved in tobacco by inserting a transgene into the chloroplast genome (Ruiz and Daniell 2005). Expression in chloroplasts of the *pha*A gene from *Acinetobacter* sp which encodes β-ketothiolase, an enzyme of the polyhydroxybutyrate pathway, resulted in a 100% male-sterile phenotype with no other phenotypic effects on plant growth and development (Ruiz and Daniell, 2005). The male sterile phenotype in this case was proposed to be caused by depletion of acetyl CoA pools in anther plastids, thereby interfering with the biosynthesis of essential fatty acids required for pollen wall development. It proved possible to partially restore fertility by growing plants for 10 days in continuous light. However, the lack of a fertility restoration strategy compatible with normal (field) growth environments renders this transgenic approach to male sterility impractical in an agricultural setting. Nevertheless, these experiments have revealed a novel alternative approach for engineering CMS-mediated hybrid seed production in the future, particularly as the number of species for which plastid transformation is feasible has recently increased (Daniell *et al.*, 2005; Kuchuk *et al.*, 2006).

5.1.2 Engineered male sterility based on CMS-determining genes

Attempts to engineer male sterility using well-characterized plant-derived CMS-determining genes has met with mixed success. The T-URF13 protein encoded by maize CMS-T mitochondria, for example, is thought to be responsible for male sterility (Dewey *et al.*, 1987) on the basis that accumulation of *T-urf13* transcripts is greatly reduced in plants restored to fertility by the R*f1* or R*f8* nuclear restorer genes (Dill *et al.*, 1997). In order to assess its utility as a general inducer of male sterility, a transgene comprising *T-urf13* fused with sequences encoding the 60 amino acid mitochondrial targeting presequence (*ß*-60) of the *ß*-ATPase 1 subunit of *Nicotiana plumbaginifolia* was introduced into the nuclear genome of tobacco (He *et al.*, 1996). The *ß*-60 targeting sequence had previously been shown to direct the import of foreign proteins into mitochondria (Boutry *et*

al., 1987; Chaumont *et al.*, 1994). Transgenic tobacco plants expressing a *β*60-T-URF13 fusion protein under the transcriptional control of the constitutive Cauliflower Mosaic Virus 35S (CaMV35S) promoter, showed sensitivity to T-toxin and methomyl but, surprisingly, did not exhibit male sterile phenotypes (Chaumont *et al.*, 1995).

Similarily, expression of the CMS-associated 25 kDa PCF protein, as a mitochondrially-targeted fusion protein, in petunia and tobacco, did not result in male sterility (Wintz *et al.*, 1995). Two fusion constructs were used, the first contained the *pcf* sequence fused with the *β*-60 mitochondrial targeting sequence under the transcriptional control of the CaMV35S promoter. In the second, the *pcf* sequence was fused to the mitochondrial targeting sequence of the *Neurospora crassa atp9* gene and expression was controlled by the tapetum-specific promoter Ta29 (Wintz *et al.*, 1995). The lack of male-sterile phenotypes, despite the accumulation of PCF fusion proteins in the mitochondria of transgenic plants, was attributed to lack of expression during a critical developmental window i.e. during the pre-meiotic stage of anther development when PCF protein is normally produced in CMS-Petunia lines.

Nevertheless, successful engineering of male sterility in transgenic plants has been reported for two CMS-determining genes. The mt genome of CMS lines of *Phaseolus vulgaris* (common bean) contains a novel open reading frame, *orf239,* which encodes a 27 kDa protein that accumulates only in reproductive tissues (Johns *et al.*, 1992). Fertility restoration with the *Fr/Fr2* nuclear restorer genes results either in complete loss of the mtDNA region containing *orf239,* or modification of *orf239* transcripts (Abad *et al.*, 1995), indicating its involvement in male sterility. Transgenic plants expressing a *β*60-ORF239 fusion protein under the control of a constitutive (sp) and a pollen specific promoter (pLat52) exhibited semi-sterile or male-sterile phenotypes at a frequency of 15% to 20%. Male-sterile plants exhibited a compacted floral head with collapsed anthers recessed well below the stigma and showed no seed set due to the absence of pollen. Semi-sterile plants had recessed anthers but an intermediate phenotype showing reduced seed set relative to fertile plants. Light microscopy staining of pollen grains was used to distinguish fertile from sterile pollen grains. Surprisingly, a control construct lacking the targeting sequence and expressing ORF239 protein in the cytoplasm also exhibited a semi-sterile or male-sterile phenotype, suggesting that

localization of ORF239 protein within mitochondria may not be necessary for male-sterility (He *et al.*, 1996).

The wheat *atp9* gene has also been used to successfully engineer male sterility in transgenic plants. Transgenes coding for fusion proteins comprising an edited or an unedited version of the ATP9 protein, fused with the mitochondrial targeting sequence of yeast *cox*IV were introduced into tobacco. Transgenic plants that expressed the unedited ATP9 fusion protein exhibited either fertile, semi-fertile, or male-sterile phenotypes (Hernould *et al.,* 1993), whereas those expressing the edited ATP9 protein were fertile. The fusion proteins were shown to be imported into mitochondria and cytological analyses indicated impaired development of the tapetal cell layer and the presence of abnormal mitochondria (Hernould *et al.,* 1998). Remarkably, it also proved possible to restore fertility in the male-sterile lines by introducing a second nuclear transgene (via cross-fertilization) coding for anti-sense *atp9* transcripts which caused a dramatic reduction in the levels of the unedited *atp9* transgene transcripts (Zabeleta *et al.,* 1996).

5.2 Chapter Aims

The aim of the work described in this chapter was to determine whether the *L. perenne* CMS-associated mitochondrial *orfC9* could be used to induce a male-sterile phenotype when expressed in transgenic tobacco plants. Since direct transformation of the mitochondrial genome is not yet possible, it was decided to express *orfC9* from a nuclear transgene and target the encoded protein into mitochondria by constructing *orfC9* in-frame gene fusions with the mitochondrial targeting sequence of the *ß-atp1* gene of *N. plumbaginifolia*.

Two different sets of transgene constructs were generated. The first set involved the design of two *ß-atp1* presequence-*orfC9* gene fusions and a control *orfC9*-only transgene that were placed under the transcriptional control of the CaMV35S promoter. In the second set of constructs, the CaMV35S promoter was replaced (in each of the *orfC9* constructs) with the anther-specific promoter *Ta29*. This chapter examines the generation and the molecular and phenotypic analysis of transgenic tobacco plants expressing six different variants of the *orfC9* gene.

5.3 Materials and Methods

5.3.1 Materials

5.3.1.1 Plasmids

pRok219 contains the Cauliflower Mosaic Virus (CaMV) 35S promoter and the nopaline synthase transcriptional terminator (NosT) in the high copy number plasmid pUC19 (T. Kavanagh, unpublished)

pBin19 is an *Agrobacterium*-derived binary vector used for plant transformation (Bevan, 1984).

5.3.1.2 Primers

5.3.1.2.1 Primers used in the assembly of the CaMV35S orfC9 construct:

The following primers were used in the amplification of the 378bp orfC9 fragment that was digested with *Bam*H1 and *Kpn*I, and cloned into the pRoK219 plasmid:

orfC9F1 5' GGG**GGATCC**ATGTTAGAAGGAGCGAAACTGGAC 3'

orfC9R2.1 5' CTC**GGTACC**TTATTCATGCGTTGTTTCTCCCTTGAG 3'

Mitochondrial targeting sequence primers:

The following primers were used in the amplification of the 180bp *ß*-60 fragment and the 270bp *ß*-90 fragment encoding the mitochondrial targeting pre-sequence of *ß-atp1*. Each digested product (with *Xba*I/*Bam*HI) was separately cloned into the pRoK219-*orfC9* vector, thereby forming pRoK219- *ß*90-*orfC9* and pRoK219-*ß*60-*orfC9*:

βatpF(90/60) 5' GGC**TCTAGA**CCATGGCTTCTCGGAGGCTTCTCGCC 3'

β atpR(90) 5' CGC**GGATCC**CCCGATCGAACCAGCGCCGGTGAA 3'

β atpR(60) 5' CGC**GGATCC**TGCCGCTGCGGAGGTAGCGTACTG 3'

Sequencing primers:

The following primers were used to confirm the assembly of the target sequence in the pRoK219-*ß90-orfC9*, pRoK219-*ß60-orfC9*, pRoK219-*orfC9* constructs containing the CaMV35S promoter.

βatpF(90/60) 5' CCATGGCTTCTCGGAGGCTTCTCGCC 3'

orfC9R1258	5' GCGAAGGTAATAATGCCCAGGC 3'
orfC9Fint	5' CAGGTTTATCCGGTTTATTCGG 3'
CamV35SF (CF)	5' GCAAGTGGATTGATGTGATATCTC 3'
orfC9R3 (OR)	5' TTATTCATGCGTTGTTTCTCCCTTG 3'
M13uni(-21)	5'TGTAAAAGGACGGCCAGT 3'
M13rev(-29)	5'CAGGAAACAGCTATGACC 3'

PCR primers:

The following primers are used to screen transgenic plants transformed with the pBin19-*ß90-orfC9*, pBin19-*ß60-orfC9* and pBin19-*orfC9* constructs containing the CaMV35S promoter.

CamV35SF (CF)	5' GCAAGTGGATTGATGTGATATCTC 3'
orfC9F.1 (OF)	5' GGG**GGATCC**ATGTTAGAAGGAGCGAAACTGGAC 3'
orfC9R3 (OR)	5' TTATTCATGCGTTGTTTCTCCCTTG 3'

RT-PCR primers:

The following primers are used to confirm the presence of transcripts in leaf tissue for the three constructs pBin19-*ß90-orfC9*, pBin19-*ß60-orfC9* and pBin19-*orfC9* containing the CaMV35S promoter.

orfC9Fint	5' CAGGTTTATCCGGTTTATTCGG 3'
orfC9intR	5' CGATATAGGGAAAATTCCAACTCGTG 3'

5.3.1.2.2 Primers used in the assembly of the Ta29 construct

Ta29 primers:

The following primers were used in the amplification of the 750bp Ta29 fragment that was digested with *Hind*III/*Xba*I and cloned into three pRoK219 constructs but this time minus the CaMV35S promoter i.e. pRoK219-*ß90-orfC9*, pRoK219-*ß60-orfC9*, pRoK219-*orfC9* :

Ta29FH	5' CGC**AAGCTT**CATGTAACAATCTGTCATTTTCG 3'
Ta29RX	5' CC**TCTAGA**AGCTAATTTCTTTAAGTAAAAACTTTG 3'

Sequencing primers:

The following primers were used to confirm the assembly of the target sequence

in the pRoK219- *ß90-orfC9*, pRoK219- *ß60-orfC9*, pRoK219- *orfC9* constructs containing the Ta29 promoter.

Ta29F1 5' CAATTGACTTGTCACTTTATGAAAG 3'

Ta29F2 5' GTAATTTCCATAACTGAAATCAGG 3'

Ta29F3 5' GTGTATGTCTTGTGCTCTATATATG 3'

Ta29R1 5' GTCACATATCCATCAAACTTCTG 3'

βatpF(90/60) 5' CCATGGCTTCTCGGAGGCTTCTCGCC 3'

orfC9R1258 5' GCGAAGGTAATAATGCCCAGGC 3'

orfC9Fint 5' CAGGTTTATCCGGTTTATTCGG 3'

orfC9R3 5' TTATTCATGCGTTGTTTCTCCCTTG 3'

M13uni(-21) 5'TGTAAAAGGACGGCCAGT 3'

M13rev(-29) 5'CAGGAAACAGCTATGACC 3'

PCR primers:

The following primers are used to screen transgenic plants transformed with the pBin19- *ß90-orfC9*, pBin19- *ß60-orfC9* and pBin19- *orfC9* constructs containing the Ta29 promoter.

Ta29FH(TF) 5' CGC**AAGCTT**CATGTAACAATCTGTCATTTTCG 3'

orfC9F.1(OF) 5'GGG**GGATCC**ATGTTAGAAGGAGCGAAACTGGAC 3'

orfC9R2.1(OR) 5' CTC**GGTACC**TTATTCATGCGTTGTTTCTCCCTTGAG 3'

Ta29 RT-PCR primers:

The following primers are used to confirm the presence of transcripts in anther tissue for the three constructs pBin19- *ß90-orfC9*, pBin19-*ß60-orfC9* and pBin19- *orfC9* containing the Ta29 promoter.

Ta29F1 5' GCAAGTTGTTGTGTATGTCTTGTGC 3'

orfC9Fint 5' CAGGTTTATCCGGTTTATTCGG 3'

orfC9intR 5' CGATATAGGGAAAATTCCAACTCGTG 3'

Extra bases that have been added to the primer for restriction digestion purposes are underlined in bold.

5.3.2 Methods

5.3.2.1 The preparation of competent *Agrobacterium* cells

Agrobacterium tumefaciens (LBA4404) was streaked for single colonies on Minimal T medium containing streptomycin (500 µg/ml) and was incubated at 28°C for 48 hr. A single colony was used to inoculate 5 ml YEB medium (0.1% yeast extract, 0.5% beef extract, 0.5% sucrose, 2 mM $MgSO_4$) containing streptomycin (500 µg/ml) using a Class I laminar air flow cabinet, and was shaken for 24 hr. at 250 rpm at 28°C. From this culture a 1 ml aliquot was removed and used to inoculate a 100 ml culture of fresh YEB-streptomycin medium, and this culture was shaken for 24 hr. at 250 rpm at 28°C. The cells were grown until they were at early to mid exponential phase (i.e. OD_{600nm} = 0.3 to 0.6). The *Agrobacterium* cells were pelleted by centrifuging for 10 min at 4000 rpm at 4°C using a Sorvall SS34 rotor. The recovered cells were gently resuspended in 10 ml cold 1 mM HEPES (pH 7.0) solution and were then re-harvested by centrifugation using the same conditions. The pelleted cells were resuspended in 5 mls of cold 1 mM HEPES (pH 7.0) containing 10% glycerol. The cells were re-pelleted as above, but this time were gently resuspended in 0.5 ml of cold 1 mM Hepes (pH 7.0) containing 10% glycerol, followed by dispensing into 100 µl amounts and storage at – 70°C.

5.3.2.2 Transformation of *Agrobacterium* by electroporation

The electroporation of *Agrobacterium* was carried out as described by Life Technologies (GIBCO BRL, InVitrogen). *Agrobacterium* cells (20 µl) were slowly allowed to thaw on ice and to this 50 - 80 ng of the plasmid construct was gently added and mixed, followed by incubation on ice for 5 to 10 min.. The cold cells were then placed into a sterile electroporation cuvette with a 1 mm gap. The sample was electroporated using a 5-6 ms pulse at the following settings: 2.4kV, 330µF. The cells were then removed and placed into a tube containing 1 ml YEB medium which was shaken at 220 rpm for 3 to 4 hr. at 28° C. The cells were spread onto YEB agar plates containing streptomycin (500 µg/ml) and kanamycin (50 µg/ml) and incubated at 28°C for 48 hr.

5.3.2.3 Transformation of *Nicotiana tabacum*

Agrobacterium-mediated nuclear transformation is based on the method described by Horsch *et al.* (1985). *Agrobacterium* transformants were grown at 28°C for 36 hr. in two separate lots of 100 ml of LB broth containing streptomycin (500 µg/ ml) and kanamycin (50 µg/ ml). The cells were pelleted by centrifugation at 4000 rpm in 50 ml amounts for 10 min at 4°C, and the pellet was resuspended in 20 ml of MS liquid medium (Murashige and Skoog, 1962). Leaves from sterile *N. tabacum* plants were wounded by generating leaf discs. These were incubated with *Agrobacterium* cells under sterile conditions for 20 to 30 min. at room temperature. The infected leaf discs were then transferred onto shoot regeneration NBM medium (MS (4.6 g/l), 0.8% agar, benzylaminopurine 1 µg/ml, α-naphthalene acetic acid 0.1 µg/ ml, B5 vitamins (inositol 100 µg/ml, thiamine 10 µg/ml, nicotinic acid 1 µg/ml, and pyridoxine-HCl 1 µg/ml) and incubated at 22°C or at room temperature for 48 hr. The leaf discs were then transferred to NBM agar medium containing carbenicillin (200 µg/ml) and kanamycin (100 µg/ml). Regenerated shoots that emerged after a period of approximately 4 weeks were excised and transferred to pots of MS selective medium. Those shoots that rooted were analysised by PCR and northern blotting.

5.3.2.4 Extraction of DNA from leaf tissue

Leaf tissue (four to six week old) was homogenized in 100 µl Extraction buffer (50 mM NaCl, 50 mM EDTA, 100 mM Tris (pH 8.0), 10 mM ß Mercaptoethanol, 100 mg/ml Proteinase K) using a sterile pestle and mortar. The homogenate was incubated at 85°C for 5 min, placed on ice for 5 min, incubated at 85°C for 5 min and replaced on ice for a further 5 min. The mixture was gently mixed and was centrifuged at 12,000 rpm for 10 min. The supernatant was removed and the pellet discarded. The supernatant was then subjected to two phenol extractions and then spin dialysed through a Sepharose CL6B mini-column to desalt the sample and remove any residual phenol (Murphy and Kavanagh, 1988).

5.3.2.5 Pollen grain counts

A single mature anther that had recently dehisced was placed into an Eppendorf tube with 300 µl Lugol potassium iodide solution. The preparation was

briefly vortexed followed by a quick centrifugation. The suspension was gently mixed, and samples of 3 to 4 drops were dispensed into the chamber wells of a haemocytometer slide. Pollen grain numbers were determined by performing counts with the use of a light microscope. Pollen grains counted were converted to number of pollen grains per anther (pga) per ml.

5.4 Results

5.4.1 Construction of gene fusions between *orfC9* and the mitochondrial targeting sequence of *ß-atp1* under the transcriptional control of the CaMV35S promoter

Sequences encoding the mitochondrial targeting pre-sequence of the ß subunit of mitochondrial F1-ATPase (*ß-atp1*) of *N. plumbaginifolia* (Boutry *et al.*, 1985) were PCR-amplified using either primer pair ßatpF/ßatpR60 or primer pair ßatpF/ßatpR90. The former amplified a 180 bp fragment (named ß60), while the latter amplified a 270 bp fragment (named ß-90). In addition to the targeting sequence, the ß-90 fragment encodes an extra 30 amino acids from the mature ß-ATP1 protein which has been shown in certain cases to improve the efficiency of the targeting pre-sequence (Chaumont *et al.*, 1994).

Gene fusions were constructed by cloning the ß-60 (or ß-90) PCR product into the CaMV35S/NosT expression cassette in pRoK219 (Figure 5.1, Figure 5.2A) using its flanking *Xba*I and *Bam*HI restriction sites. In a second cloning step, *orfC9* was PCR-amplified from the pCMS60 plasmid and was introduced downstream of ß-60 (or ß-90) using the *Bam*HI and *Kpn*I restriction sites (Figure 5.1, Figure 5.2B). This resulted in an in-frame fusion of the *orfC9* gene to ß-60 (or ß-90) under the transcriptional control of the CaMV35S promoter. The resulting constructs were called 35S-ß60-orfC9 and 35S-ß60-orfC9, respectively. A control *orfC9* construct lacking the mitochondrial targeting pre-sequence, 35S-orfC9, was also made (Figure 5.1). Confirmation of construct inserts by restriction digestion is shown in Figure 5.2C. The integrity of the various constructs was also confirmed by sequencing using the CaMV35S promoter primer (CaMV35SF), targeting pre-sequence primers (ß-90/ ß-60), *orfC9* primers (orfC9R) and universal primers (Materials 5.2). The entire sequence of the *Hind*III/*Eco*RI region in each of the three Rok219 constructs is shown in Figure 5.3.

Each of the transgene expression cassettes was excised from pRok219 by restriction digestion at the flanking *Hind*III and *Eco*RI sites e.g. a 1.695 kb 35S-ß60-orfC9 *Hind*III-*Eco*RI restriction fragment was ligated into the *Agrobacterium* binary vector pBin19 to give the plant transformation vector pBin19-35S-ß60-orfC9. The pBin19-35S-ß90-orfC9 and pBin19-35S-orfC9 vectors were

constructed in the same way using *Hind*III-*Eco*RI fragments from the two respective pRoK219 constructs (Figure 5.1). Confirmation of the *Hind*III/*Eco*RI insert in each of the pBin19 constructs was also carried out by sequencing. The appropriate *Hind*III/*Eco*RI fragment for each of the three constructs was sub-cloned back into the high copy pUC19 plasmid which was then sequenced using the same primers used in the sequencing of the three original pRok219 constructs. The sequence of relevant pBin19 *Hind*III/*Eco*RI transgene cassette fragments was identical with the sequence of the pRok219 constructs (Figure 5.3).

The three pBin19 constructs were transformed into the *Agrobacterium* strain LBAA4404 by electroporation and colonies selected on MinT agar plates containing streptomycin (500 μg/ml) and kanamycin (50 μg/ml). The integrity of the constructs in *Agrobacterium* was verified by re-transforming them back into *E. coli* DH5-α, with selection on LB Kanamycin (50 μg/ml) medium. Minipreparation plasmid DNA was isolated and the structure of the cassettes verified by restriction digestion with *Hind*III and *Eco*RI (Figure 5.2D).

Agrobacterium clones containing the various constructs were used to transform *N. tabacum* (var Samsun NN) by means of the leaf disc transformation method (Horsch *et al.*, 1985).

5.4.1.1 Screening of the transgenic plants by PCR

The transformed leaf discs of *N. tabacum* were selected on MS-NBM shoot regeneration medium containing 100 μg/ml kanamycin and 200 μg/ml carbenicillin. Shoots were excised and placed on MS medium containing kanamycin (100 μg/ml) and carbenicillin. (200 μg/ml). After 2-3 weeks, DNA samples from putatively transformed shoots that showed evidence of rooting were tested by PCR analysis using the CaMV35SF (CF) primer and orfC9R2.1/orfC9R3 (OR) primer pair. The location of the primers used for PCR screening (see 5.2.1.Materials) and the corresponding size fragments is shown in Figure 5.3 and 5.4. Unique single PCR products of 775 bp, 685 bp and 503 bp were amplified from transformants containing the constructs 35S-*ß*90-orfC9, 35S-*ß*60-orfC9 and 35S-orfC9, respectively (Figure 5.5). The PCR fragment for each construct was further confirmed by direct sequencing. In summary,

transformation with the S35-ß60-orfC9 construct yielded 15 kanamycin-resistant (Kan-R) shoots of which 14 were PCR positive, while the
35S-orfC9 construct yielded 11 Kan-R shoots of which 10 proved positive by PCR. The S35-ß90-orfC9 construct gave 13 Kan-R shoots of which 11 proved to be positive for the transgene by PCR (Figure 5.5). A second set of PCR amplifications was performed on the same putatively positive transformants using the ORFC9F(OF)/ORFC9R(OR) primer pair (Figure 5.4). These positively transformed plants (T1 to T30) contained an expected fragment of 378 bp (Figure 5.4, Figure 5.6). The sequence of the orfC9 PCR fragment was confirmed by using the primer set (orfC9F (OF)and orfC9R(OR)).

5.4.1.2 Northern blot detection of orfC9 transcripts in To transgenic lines

A total of 30 different transgene-positive lines representative of the three orfC9 constructs were rooted and propagated in a greenhouse under natural light conditions at 25°C. Total RNA was extracted from leaf tissue (four weeks old) for 25 of the 30 lines and 10 µg of total RNA was electrophoresed for each line on a 1% denaturing agarose gel and northern blotted by capillary action onto a standard nylon membrane (RPN2013). The blot was then hybridized with a randomly-primed ^{32}P-labelled orfC9 PCR fragment (378bp).

A mono-cistronic transcript was detected for each of the three different constructs: transformants containing 35S-ß90-orfC9 gave a transcript of 1.0 kb, 35S-ß60-orfC9 a transcript of 0.9 kb, and 35S-orfC9 a transcript of 0.7 kb. Although equal amounts of total RNA were probed for each line, the transcript intensities varied between different lines (Figure 5.7A).

Transcripts were readily detected (ca. at 12 to 18 hr. exposure) in the 35S-ß90-orfC9 transformed lines T1, T2, T3, T4, T5, T7; in the 35S-ß60-orfC9 lines T11, T14, T15, T18; and also in 35S-orfC9 lines T22, T24 and T25. However, faint transcripts were visible in line T10 (a 35S-ß90-orfC9 line), T12, T16 (both 35S-Sß60-orfC9 lines), T27 (a 35S-orfC9 lines) and no transcripts were visible for T6 and T13 (both 35S-ß90-orfC9 lines) as shown in Figure 5.7A.

5.4.1.3 RT-PCR detection of orfC9 transcripts in total RNA extracted from anthers of To transgenic lines

Total RNA was also extracted from 200-300 mg of anther tissue of the 35S-*ß*90-orfC9 lines T2, T3 and T4, yielding 4-5 μg of total RNA (Figure 5.7B). When total RNA (ca. 5μg) from leaf tissue of the same lines was probed with orfC9, a single transcript of 1.0 kb was seen in each line. Line T4 showed the greatest transcript abundance followed by lines T3 and T2. (Figure 5.7B). However, northern blot analysis of total RNA (2 μg) from anther tissue of lines T2, T3 and T4 gave no evidence of orfC9 transcripts, suggesting that the constitutive CaMV35S promoter may not be as active in anther tissue as in leaf tissue (Figure 5.7B).

However, randomly primed single-stranded cDNA was generated from total anther RNA of lines T3 and T4, and of the non-transformed (Nt) line, using Supercript II RTase. A PCR product of 260 bp (Figure 5.8) was detected when using the orfC9Fint/orfC9Rint primer set (for orfC9 gene amplification), in samples of the T3 and T4 lines that had been reverse transcribed. No PCR product was present in the samples of the same lines that had not been reverse transcribed, indicating that one can rule out the likelihood of contamination by genomic DNA.

5.4.1.4 Northern blot detection of orfC9 transcripts in total RNA extracted from leaf tissue of T1 transgenic lines

A series of T0 transformed lines that gave a single transcript when probed with orfC9 (Figure 5.7A) had their T1 generation (selfed) lines examined further by northern blotting, with a view to determining how stable transcript expression was in their progeny. Total RNA (5-10 μg) was extracted from T1 leaf tissue of lines T3, T4, T5, T7 (for the S35-*ß*90-orfC9 construct), T14, T15, T20 (for S35-*ß*60-orfC9) and T21, T27, T29 (for the 35S-orfC9 construct). The correct size transcript was detected in the T1 generation for each of these lines (Figure 5.9). The transcript was abundant for all of the T1 lines except for the T21 line (Figure 5.9), which interestingly also showed low transcript abundance in the T0 generation (Figure 5.7A).

5.4.1.5 Analysis of sterility based on flower morphology, pollen production and seed set

The regenerated T0 plants S35-*ß*90-orfC9, S35-*ß*60-orfC9, 35S-orfC9 were grown to flowering and were assessed for sterility on the basis of flower morphology, pollen grain counts per anther (pga) and seed set. The male sterility phenotype was classified as either sterile or semi-sterile. Sterile plants showed defects in anther development and anthers that were recessed below the level of the stigma, and by fruit capsules containing little or no seeds. Semi-sterile plants had recessed anthers and produced fruit capsules with a reduced seed set (He *et al.*, 1996).

Amongst the ten T0 35S-*ß*90-orfC9 transgenic lines expressing the *ß90orfC9* transcripts, only 20% of the plants displayed a semi-sterile phenotype. The T3 and T4 transgenic lines produced recessed anthers relative to the stigma, and the stamens were shorter than in the wild type plants (Figure 5 10A, B). It was also noticed that in some of the flowers of the line T3 the anthers failed to dehisce (Figure 5.10C). In addition, for the lines T3 and T4, the size of the fruit capsules and seed set was significantly reduced, compared with other transformed lines (T1, T5, T7), and non-transformed lines showing a male fertile phenotype (Figure 5.10D).

The pollen grain number per anther (pga) was counted with the use of a haemocytometer (Table 5.1). Values of 4.6 x 10^4 and 13.8 x 10^4 pollen grains per anther were obtained for lines T3 and T4, respectively (Table 5.1). The number of pollen grains per anther in the remaining 35S-*ß*90-orfC9 lines (T1, T2, T5 to T10) was 17.2 x 10^4, whereas in the non-transformed line 10 flowers gave a value of 22 x 10^4 pollen grains per anther. The number of pollen grains per anther was approx 4-fold lower for the T3 line but was only 1.5 fold lower for the T4 line in comparison to non-transformed (Nt) line (Table 5.1). Sterility of pollen grains was assessed by staining for cell contents with Lugol solution (potassium iodide). For lines T3 and T4, both of which produced recessed anthers, 59.5% and 40%, respectively, of the pollen grains were sterile as evidenced by their lack of contents and their collapsed, elliptical exines (Table 5.1). Whereas, pollen grains from non-recessed anthers (Figure 5.11C (i)) produced by non-transformed (male fertile) lines was stained fully purple, with 95% of the cells having a round shape typical of fertile pollen cells (Table 5.1; Figure 5.11C (ii)). Similarily, pollen

grains from non-recessed anthers in remaining 35S-ß90-orfC9 transformed lines (Figure 5.11D(i)), stained fully purple, with 92.9% of the pollen cells (an average value for 10 flowers over 8 different plant lines), appearing round in shape (Table 5.1; Figure 5.11D (ii)). The staining of pollen grains with Lugol solution was used here only as an indicator of fertility; it cannot confirm pollen viability.

Self-fertilisation of T3 and T4 plants resulted in reduced seed set as shown in Table 5.1. The T3 line gave a mean seed weight per fruit capsule of 32.5 mg (over 10 capsules) for two transgenic plants. The T4 line failed to produce any seeds in the majority of fruit capsules but gave a reduced seed set of 12.5 mg (over 10 fruit capsules) for two plants. This was in contrast to a mean seed set (per fruit capsule) of 126 mg for the remaining 8 lines of the 35S-*ß*90-orfC9 construct (over 10 fruit capsules per line), all of which showed a male fertile phenotype (Table 5.1). The non-transformed (Nt) or wild type seed set was assessed for 4 plants, and all of the plants as expected showed a male-fertile phenotype, with an average mean seed set of 140 mg per fruit capsule.

To investigate the possibility that plants were semi-sterile due to low numbers of pollen grains or that the pollen was non-viable, a total of 12 hand-emasculated flowers from two non-transformed lines were cross-pollinated with pollen collected from anthers of the T3 and T4 transgenic lines (T_0 generation plants). The crosses resulted in either a reduced or no seed set (Figure 5.12A): 15 - 40 mg/fruit capsule for the cross T3 x Nt and the line T4 x Nt. These seeds were then germinated on MS medium containing kanamycin (500 µg/ml) as shown in Figure 5.12C.

It is important to note that shorter stamens could be due to the failure of the cells to properly elongate in the central and upper regions of the filament. Therefore, in order to investigate the possibility that sterility in the T3 and T4 lines was due to female rather than male infertility, T3 and T4 plants were fertilised (ca. 10/15 flowers (5 flowers from separate plants)) with pollen from fertile non-transformed wild-type lines (Nt) (Figure 5.12B). The resulting fruit capsules were normal in size and appearance and produced a normal seed set of 140 mg mean weight. The seeds from the cross Nt x T3 and the cross Nt x T4 were able to germinate and develop normally on kanamycin (500 µg/ml) MS medium (Figure 5.12C).

Amongst the 10 T_0 35S-ß60-orfC9 and 35S-orfC9 transgenic lines expressing the *orfC9* transcripts, neither a sterile nor a semi-sterile phenotype was observed. The anthers for all these plant lines were located well above the stigma, typical of fertile non-transformed plants (Figure 5.13A, B(i)). The pollen grains were round and fully stained purple (Figure 5.13A, B (ii)), and the fruit capsules were normal in size (data not shown). The fruit capsules produced a normal seed set of 153.3 mg for 35S-ß60-orfC9 transformant and 154 mg for the 35S-orfC9 transformant lines (Table 5.1). The number of pollen cells for the 35S-ß60-orfC9 construct lines was 14.4 x 10^4 cells per anther for 10 flowers over 10 plants which was slightly less compared with fertile non-transformed anthers. The average number of fertile pollen grains per flower for the 35S-ß60-orfC9 transformant was 89.7% (Table 5.1). Likewise, the number of pollen cells for the 35S-orfC9 transformant lines was 15.8 x 10^4 cells per anther for 10 flowers (over 10 plants), being slightly less in comparison with fertile non-transformed anthers. The average number of fertile pollen grains per flower for the 35S-orfC9 construct was 92.2% (Table 5.1).

It appears from these findings that semi-sterility is associated with a partial reduction in pollen number and the percentage level of sterile pollen cells per anther. This suggests that *ß90-orfC9* expression from a CaMV35S promoter may disrupt pollen development in tobacco.

5.4.1.6 Analysis of the seedlings phenotype in the different transgenic lines.

Seeds were harvested and collected from plants grown to flowering in the greenhouse for the 30 different transgenic lines. Seeds from a series of these transformed lines along with non-transformed seeds were allowed to germinate on MS-agar plates containing kanamycin (500 µg/ml) and were scored after a 3 - 4 week growth period. The T_1 generation transgenic progeny segregated as green kanamycin-resistant (Kan-R) and bleached kanamycin-sensitive (Kan-S) seedlings (Table 5.2). The control non-transformed wild-type seedlings were fully bleached. The segregating transgenic lines T5 and T21, and possibly the lines T2 and T25 likely contain a single copy of the transgene and therefore show a typical Mendelian segregation ratio of 3:1. The line T27 gave a segregation ratio of 9.6:1, indicating two copies of the transgene. The lines T3, T4, T7, T14 and T15 showed

no segregation, indicating that these lines might contain multiple transgene insertions (Table 5.2).

Although at the T0 generation, the lines T3 and T4 showed a semi-sterile phenotype whereas the T1, T22, T24 and T29 lines were fertile, seed collected from each line germinated normally on MS-Kan medium as shown in Figure 5.14. After a period of 5-6 weeks, seedlings from each line and a fertile (transgenic) positive control line were potted in compost and allowed to grow to flowering. Flower produced by these T1 generation plants showed the same phenotype as the To parent plant in terms of the position of the anthers. The anthers for the lines T3 and T4 were significantly recessed beneath the stigma typical of the flower morphology associated with a semi-sterile phenotype (Figure 5.15A, B). The lines T22, T24 and T29 had their anthers significantly above the stigma as observed for the fertile phenotype. The fruit capsules were normal in size with a normal seed set value (98 mg per fruit capsule (Figure 5. 15 A, B).

Seeds were harvested from these T1 plants and were plated onto MS-Kan medium as shown in Figure 5.14. Seedlings of the T2 generation T3 and T4 lines were further cultured on fresh MS-Kan for a further 5 - 6 weeks, were potted in soil and grown to flowering. The flowers produced by each line were again semi-sterile with recessed anthers (Figure 5.16 A, B). The fruit capsules were of reduced size with a reduced seed set. The T3 line gave a mean weight of 13.33 mg per fruit capsule and the T4 line yielded a mean weight of 19.21 mg per fruit capsule (Table 5.3). But when the stigma of the T3 and T4 flowers was cross-fertilised with fertile pollen from a non-transformed line (Nt) (Figure 5.16C), a normal size fruit capsule was produced with a seed set of 130 mg seed/capsule (data not shown).

5.4.1.7 Western blot investigation of ORFC9 protein expression in transgenic lines

Total protein was extracted from T0 generation leaf tissue of the following lines: T2, T3, T4, T7, T8 (containing the 35S-*β*90-orfC9 construct); T11, T13, T14, T15, T16, T18 (containing the 35S-*β*60-orfC9 construct); and T21, T23, T27, T28 (containing the 35S-orfC9 construct). The non-transformed line (Nt) was used as a negative control. Samples of total protein for each line was electrophoresed on a 14% SDS-PAG along with insoluble recombinant ORFC9

protein (Ip) as a positive control (Chapter 3) as shown in Figure 5.17. The fractionated proteins were then transferred to a nitrocellulose membrane which was probed using the previously described anti-ORFC9 polyclonal antibody (Figure 5.17). The anti-orfC9 antibody, which readily detected the control recombinant ORFC9 protein, did not however reveal any bands in the plant-derived protein samples. This suggests that the ORFC9 protein does not accumulate in leaf tissue of transgenic lines at a level sufficient for detection by the polyclonal antibody.

5.4.2 Use of the Ta29 tapetum-specific promoter to drive *orfC9* expression

The Ta29 promoter was used to drive orfC9 expression in three constructs analogous to those described above (5.4.1.1) and illustrated in Figure 5.18. The Ta29 promoter was amplified by PCR from total DNA of *N. tabacum* using the primer pair Ta29FH and Ta29RX (Figure 5.19A) which include, respectively, *Hind*III and *Xba*I restriction sites. The inclusion of these sites facilitated replacement of the CaMV35S promoter in the constructs pRok219-35S-*ß*90-orfC9, pRok219-35S-*ß*60-orfC9 and pRok219-35S-orfC9 with the amplified Ta29 promoter fragment following digestion with *Hind*III and *Xba*I (Figure 5.18, 5.19B). The pRok219-*ß*90-orfC9 *Hind*III/*Xba*I vector fragment was gel-purified and was then ligated with the Ta29 *Hind*III/*Xba*I-digested promoter fragment thereby yielding the construct pRok219-Ta29-*ß*90orfC9. The same approach was used to generate the pRok219-Ta29-ß60-orfC9 and pRok219-Ta29-orfC9 constructs (Figure 5.18). The presence of the Ta29 promoter fragment in the three pRok219 constructs was verified with the restriction *Hind*III/*Xba*I digests (Figure 5.19C). Furthermore, the entire insert containing the promoter, mitochondrial targeting sequence (*ß*-90, *ß*-60) and *orfC9* was confirmed by sequencing using the appropriate primers (Materials section 5.2). The entire sequence of the *Hind*III/*Eco*RI region 1.67 kb, 1.58 kb, 1.4 kb in each of the pRok219 constructs is shown in Figure 5.20.

The fragments Ta29-*ß*90-orfC9, Ta29-*ß*60-orfC9, Ta29-orfC9 (including the 3' NosT terminator) were each sub-cloned into the low copy Agrobacterium binary vector pBin19 at the *Hind*III/*Eco*RI sites to give: pBin19-Ta29-*ß*90-orfC9, pBin19-Ta29-*ß*60-orfC9 and pBin19-Ta29-orfC9 and each in turn was transformed into the *Agrobacterium* strain LBAA4404 by electroporation.

The integrity of the three pBin19 constructs in *Agrobacterium* was confirmed by re-transforming them back into *E.coli*, preparing miniprep plasmid DNA, and restriction digestion with the enzymes *Hind*III and *Eco*RI to confirm the presence of the inserts (Figure 5.19D). *Agrobacterium* colonies containing the construct of interest were used to transform *N. tabacum* (Samsun NN) by means of the leaf disc transformation method (Horsch *et al.*, 1986).

5.4.2.1 Screening of the Ta29 transgenic plants by PCR

Putatively transformed shoots were screened by PCR analysis using the primer pair Ta29F and orfC9R. The Ta29-*ß*90-orfC9 construct generated 12 Kan-R shoots of which 10 proved to be positive by PCR, the Ta29-*ß*60-orfC9 construct yielded 13 Kan-R shoots of which 10 were positive, and the Ta29-orfC9 construct yielded 12 Kan-R shoots of which 10 proved positive by PCR.

The position of the primers used for PCR amplification of the fragments is shown in Figure 5.21. Diagnostic PCR fragments of 1.38 kb, 1.2 kb and 1.11 kb were amplified respectively, from Ta29-*ß*90-orfC9, Ta29-*ß*60-orfC9 and Ta29-orfC9 and their identity confirmed by direct sequencing (Figure 5.21, Figure 5.22A, B).

The primer pair ORFC9F/ORFC9R, which amplifies a 378 bp fragment of orfC9, was also used to screen the putatively positive transformants: lines T1 - T30 contained the expected fragment (Figure 5.21, Figure 5.23). The orfC9 PCR fragment was also confirmed by direct sequencing.

5.4.2.2 Analysis of *orfC9* transcripts in anther tissue of the To transgenic lines by RT-PCR

Total RNA was extracted from the To leaf tissue from several independent transformant lines of the three different constructs and northern blot analysis confirmed the expected lack of expression of orfC9 in the leaf tissue (data not shown).

Since the Ta29 promoter is active only in the tapetal cell layer, it was decided to use RT-PCR to investigate the presence of orfC9 transcripts in the anther tissue for the three different constructs. Total RNA was also extracted from To anther tissue at stages 3 - 6 of anther development, at which point the flower buds are still closed, the anthers are non-dehiscent, and tapetal cell degradation

has not yet started (Koltunow *et al.*, 1990). Total RNA was extracted from ca. 200-300 mg of anther tissue of the transgenic lines T3, T5, T13, T17, T18, T19 and T29, yielding 4-5 μg of total RNA.

RT-PCR was performed as described above. A PCR product of 260 bp was amplified when using the primer set orfC9Fint/orfC9Rint in the lines T3 and T5 (containing the Ta29-*ß*90-orfC9 transgene); T13, T17, T18 and T19 (containing the Ta29-*ß*60-orfC9 transgene); and T29 (containing the Ta29-orfC9 transgene) (Figure 5.24 a). No fragments were amplified in samples of the same lines that were non-reverse transcribed. The line T3CL which refers to cDNA from leaf RNA of a transgenic line transformed with a CaMV35S construct was used as a positive control.

Given the fact that orfC9 transcripts were detected by RT-PCR, attempts were then made to identify transcripts by Northern blotting, remembering of course that we are dealing with small amounts of anther tissue/buds (Stages 3 to 6). Total RNA extracted from anther tissue and leaf tissue of the Ta29 lines T5, T17, T18, T29 was probed with *orfC9* (Figure 5.25 A, B). A negative control was set up by probing total RNA extracted from leaf tissue of the Ta29 lines T5, T17, T18 and T29 in Figure 5.25B, and as expected no transcripts were detected in leaf tissue in any of these lines (Figure 5.24).

Furthermore regarding anther tissue, no evidence of an *orfC9* transcript was obtained for any of the Ta29 lines, suggesting that the orfC9 transcript does not accumulate to high enough abundance in anther tissue for detection by Northern blotting (Figure 5.25B).

5.4.2.3 Analysis of sterility by flower morphology

The analysis of flower morphology was carried out as described above (Section 5.4.1). Examination of flowers of transgenic lines representative of the three constructs did not reveal sterile or semi-sterile phenotypes, in that none of the anthers were recessed below/beneath the stigma (Figure 5.26A (i),(ii)).

Anthers from 30 transgenic lines also did not show a reduction in pollen grain number as determined using a haemocytometer compared with a non-transformed (Nt) line (Table 5.4). A value of 2.52 $x10^5$ pollen grains per anther (pga) was recorded for the non-transformed line, compared with 2.2 $x10^5$ pga for

Ta29-β90-orfC9 transformant, 3.24 x10^5 pga for Ta29-β60-orfC9 transformants, and 1 97 x10^5 pga for Ta29-orfC9 transformants. Staining with Lugol solution failed to detect any sterile pollen grains lacking cell contents. Hence, fertility levels of 91% were recorded for Ta29-β90-orfC9 lines, 93% for Ta29-β60-orfC9 lines, and 96% for Ta29-orfC9 lines as shown in Table 5.4.

The fruit capsules from all of the transformed lines were of the same size as fruit capsules from non-transformed lines (Figure 5.26B (i), (ii)), and likewise the mean seed set mass (mg) was similar in value for both the transformed and the non-transformed lines (Table 5.4). The Ta29-β90-orfC9 transformant gave a seed set value of 122 mg per capsule compared to 135 mg for the non-transformant (Nt), whereas the Ta29-β60-orfC9 and Ta29-orfC9 transformants gave values of 111.6 mg and 138.8 mg, respectively. The seeds that were obtained from the fruit capsules for the To plants in the chosen transgenic lines T5, T17, T18, T29 were shown to be capable of germinating on MS-Kan medium (Figure 5.27).

5.5 Discussion

The chimeric gene *orfC9* is associated with CMS in *L. perenne* (Kiang and Kavanagh, 1996). But it is not yet known whether *orfC9* plays a causative role in the CMS phenotype. Attempts were made in this study to investigate whether a fusion protein comprising the mitochondrial targeting sequence of *ß*-ATP1 fused with ORFC9 could induce a male-sterile phenotype in tobacco when expressed from a nuclear transgene. Transgenes were produced in which *orfC9* was fused with sequences coding for the *ß*-60 (60 amino acids) or *ß*-90 (90 amino acids) targeting sequences. These had previously been shown to direct the import of GUS and CAT (chloramphenicol acetyltransferase) into the mitochondria of transgenic plants (Boutry *et al.*, 1987; Chaumont *et al.*, 1994). The efficiency of mitochondrial import of a GUS fusion protein was increased when GUS was fused with the longer *ß*-90 sequence (Chaumont *et al.*, 1994); both targeting sequences are predicted to be cleaved at the same site i.e. Tyr-54, Ala-55 (Chaumont *et al.*, 1994).

Transgenic *N. tabacum* plants were generated by *Agrobacterium*-mediated transformation of leaf discs using 6 *orfC*9 constructs: three with the CaMV35S and three with the Ta29 promoter. For the CaMV35S constructs a total of 39 independent transformants were obtained of which 35 were positive for the transgene by PCR analysis and 30 were then assessed for male sterility. A total of 37 independent transformants were obtained for the Ta29 promoter constructs of which 30 tested positive by PCR and were assessed phenotypically.

Only 20% of T_0 generation 35S-*ß*90-orfC9 transgenic plants exhibited a floral morphology indicative of male sterility, with the anthers recessed beneath the stigma and a high proportion of pollen grains lacking cell contents. This male sterile phenotype persisted in the T_1 and T_2 generation plants. Interestingly a similar percentage was reported by previous workers, investigating expression of the *P. vulgaris* CMS *orf239* in tobacco (He *et al.*, 1996). Here, neither the To nor T_1 generation gave 100% sterility but rather a 15 - 24% sterility across the different constructs, both with and without the *ß*-60 presequence. However the ORF239 protein is unique in that it is not located in mitochondria in either CMS *P. vulgaris* or transgenic tobacco plants (He *et al.*, 1996). Rather it functions outside of the mitochondrion, being active in the pollen cell walls and in the

callose layer of the microspore tetrads. Thus, it does not require targeting to the mitochondrion to induce male sterility in tobacco. In *N. tabacum*, no male sterile phenotype was identified in the lines expressing the 35S-orfC9, 35S-ß60-orfC9, Ta29-orfC9, Ta29-ß60-orfC9 and Ta29-ß90-orfC9 transgenes. It may be that the extra 36 aa-residues of the ß-90 targeting sequence in 35S-ß90-orfC9-transformants, relative to ß-60 sequence may enhance the efficiency with which ORFC9 protein is targeted to mitochondria, thus explaining why a sterile phenotype was only observed among the 35S-ß90-orfC9 transformants.

Expression of the maize T-URF13 protein with or without a targeting sequence failed to induce male sterility in tobacco (Chaumont *et al.*, 1995). Sensitivity to the insecticide methomyl was detected in 6.25% of the ß60-T-urf13 and 50% of the T-urf13 fertile transformants, and was attributed to the *T-urf13* gene. However, a 10% semi-sterile, and a 10% sterile phenotype was also detected in ß60-T-urf13 transformants, but these did not show methomyl sensitivity. For this reason it was concluded from analysis of their progeny that the sterility was not linked to the *T-urf13* gene but was possibly due to an artefact of *in vitro* cell culture. Although it is possible that the semi-sterility observed in 2 of the 10 transformants containing the 35S-ß90-orfC9 transgene (lines T3 and T4 lines) was caused by factors other than the transgene, this seems unlikely, because semi-sterility was not observed in any of the 50 transgenic lines expressing the 35S-orfC9, 35S-ß60-orfC9, Ta29-orfC9, Ta29-ß60-orfC9 or Ta29-ß90-orfC9 transgenes. It is also possible that the semi-sterile phenotype observed in lines T3 and T4 was caused by insertion of the 35S-ß90-orfC9 transgene into an endogenous plant gene. Once again this is also unlikely because most tobacco genomic DNA is non-coding and phenotypes caused by transgene insertion into endogenous plant genes are rarely observed.

Why might the level of observed sterile phenotypes be so low? To address this point one may initially look at the choice of promoter. The CaMV35S and Ta29 promoters are relatively inactive in sporogenous cells at the pre-meiotic stage (Wintz *et al.*, 1995), showing greater activity in post-meiotic cells or at the gametophytic stage. In *L. perenne,* expression of ORFC9 protein in sufficient quantities at a precise stage in pollen development may be a pre-requisite for the induction of male sterility. Examination of several fertile lines containing the 35S-ß90-orfC9, 35S-ß60-orfC9, 35S-orfC9 transgenes showed a relatively high level

of *orfC9* transcripts, but only in leaf tissue. Northern blot analysis showed expression of *orfC9* transcripts in leaf tissue for each of the CaMV35S lines (25 plants). The majority of the lines examined produced a male fertile phenotype (except for T3 and T4), even though abundant transcripts were detected in leaf tissue. This might be due primarily to inability of the CaMV35S promoter to direct expression of the *orfC9* gene at the correct developmental stage in anthers.

The CaMV35S promoter is constitutively active in leaf tissue, but there are conflicting reports regarding its activity in tapetal cells. Some workers report weak activity in anther tissue of transgenic petunia and tobacco plants (Wintz *et al.*, 1995; Duroc *et al.*, 2006). However, GUS activity has been detected in the tapetal cell layer when GUS expression was controlled by CaMV35S (Hernould *et al.*, 1998). In my study northern blotting failed to detect *orfC9* transcripts in total RNA extracted from anther tissue of the fertile line T2 and the semi-sterile lines T3 and T4. However, RT-PCR was able to detect orfC9 transcripts in anthers of T3 and T4 plants. This suggests that CaMV35S activity and in consequence *orfC9* transcript levels are low in anther tissues. This also likely means that the putative ORFC9 protein is expressed at low levels, if at all, in anthers and that the level may not be sufficient to induce male sterility in most transgenic lines.

Regarding the tapetum-specific promoter Ta29, I predicted that using such a promoter might increase the frequency of male-sterile phenotypes by directing developmentally appropriate expression of *orfC9* in the correct tissue. However, although the anthers of Ta29-controlled lines examined by RT-PCR showed evidence of *orfC9* transcripts, no transgenic lines exhibited a male-sterile phenotype. Similar studies have shown that attempts to use this promoter to drive expression of the petunia CMS-determining *pcf* gene in tobacco failed to produce male-sterile plants, even though the 25 kDa PCF protein was expressed in tobacco (Wintz *et al.*, 1995).

The nature of the involvement of the *orfC9* transgene in the male sterility phenotype observed in the CaMVS35 transformant lines T3 and T4 is still unclear and one cannot yet discount the possibility that the semi-sterile phenotype might either be a result of *orfC9* transgene interference in the nuclear genome, or the result of *in vitro* cell culture inducing a genome modification that fortuitously resulted in male sterility. A similar conclusion was reported for the induced male-

sterile phenotypes observed in tobacco plants expressing the T-URF13 (Chaumont *et al.*, 1994).

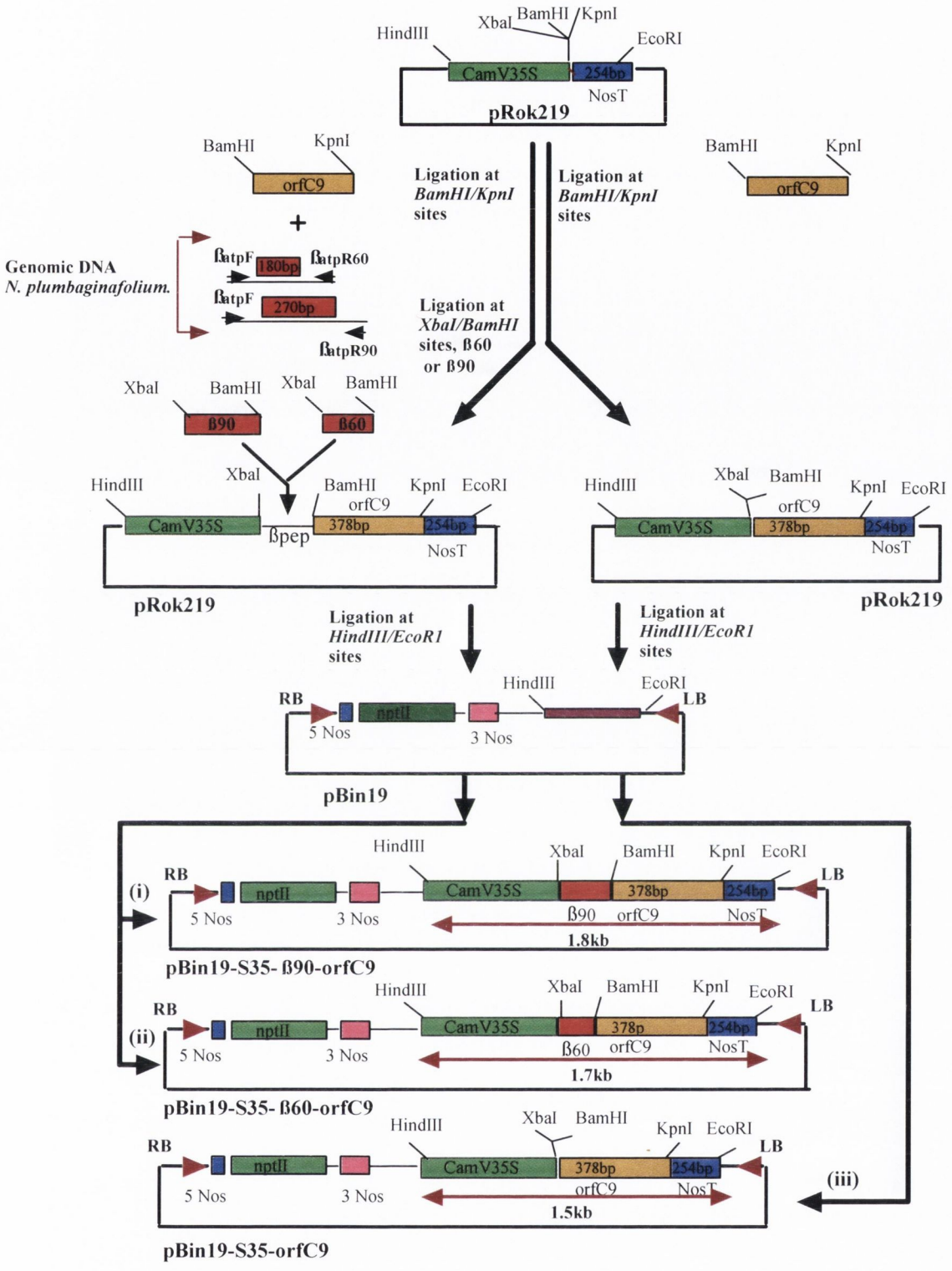


Figure 5.1 Schematic construction of the pBin19-35S-βsubunit-orfC9 and pBin19-35S-orfC9 constructs.

Primers ßatpF, ß90atpR, ß60atpR used in the amplification of the transit ß-peptide sequences from *N. plumbaginafolia* The CamV35S transcriptional promoter, the transit peptide encoding the *ß*-90 and *ß*-60 NH_2-terminal amino acid residues of the ATPase **ß**-precursor and orfC9 gene encoding the putative CMS protein ORFC9 constitute three pBin19 constructs.. (i) pBin19-35S-*ß*90-orfC9 (ii) pBin19-35S-*ß*60-orfC9 (iii) pBin19-35S-orfC9 (diagram not to scale).

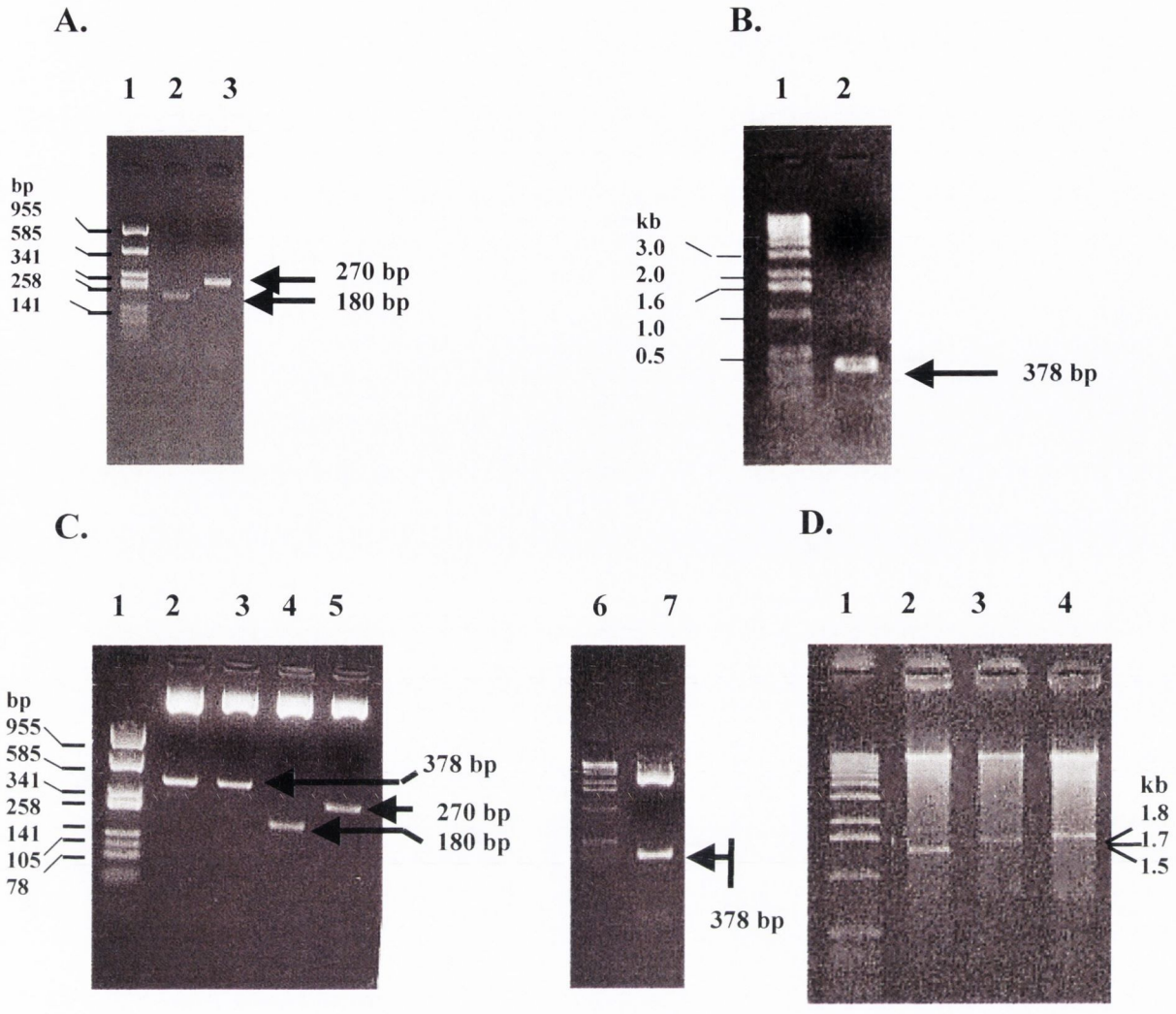
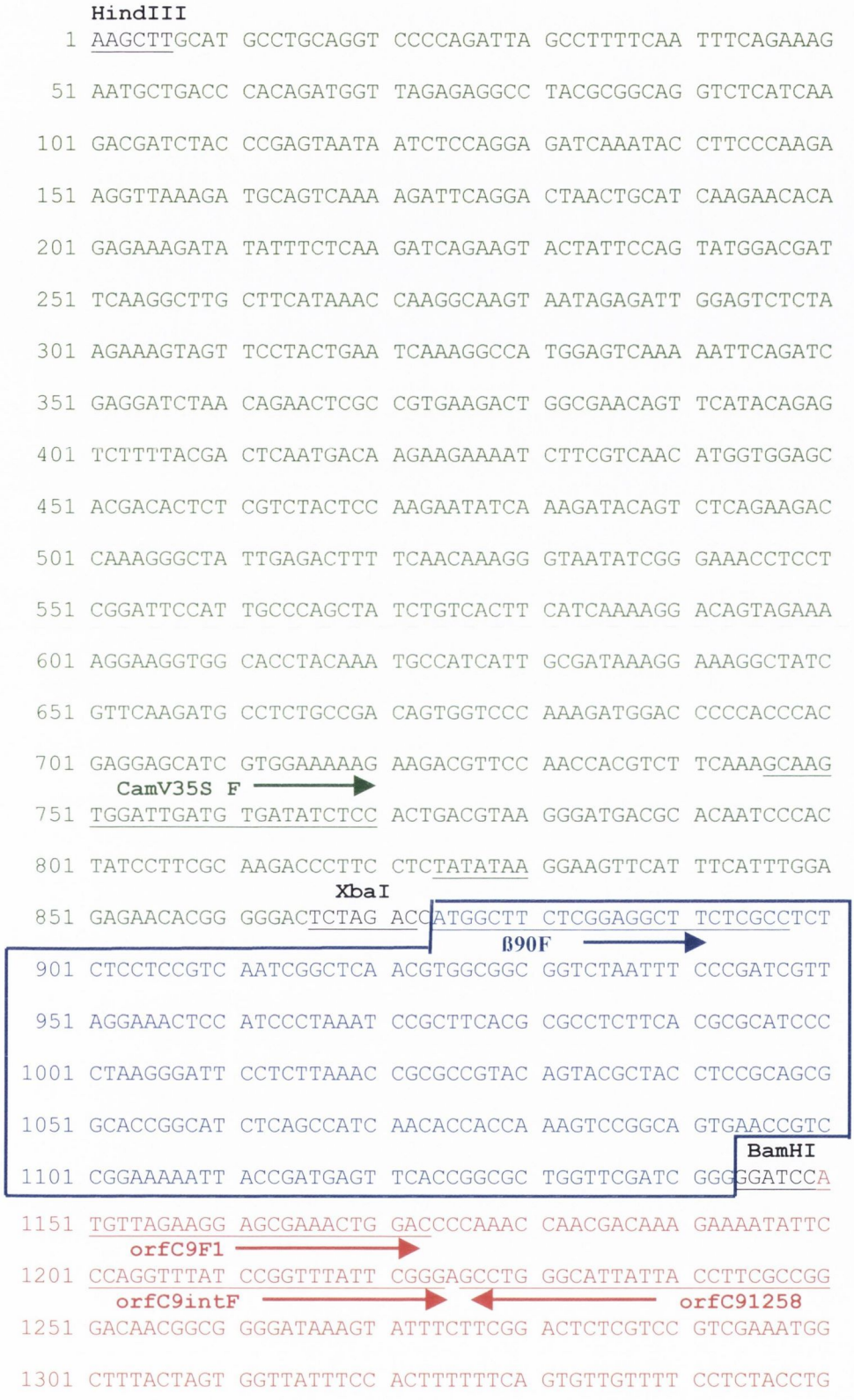


Figure 5.2 PCR products and restriction digests confirming construction of pBin19-35S- βsubunit-orfC9, pBin19-35S-orfC9 constructs.

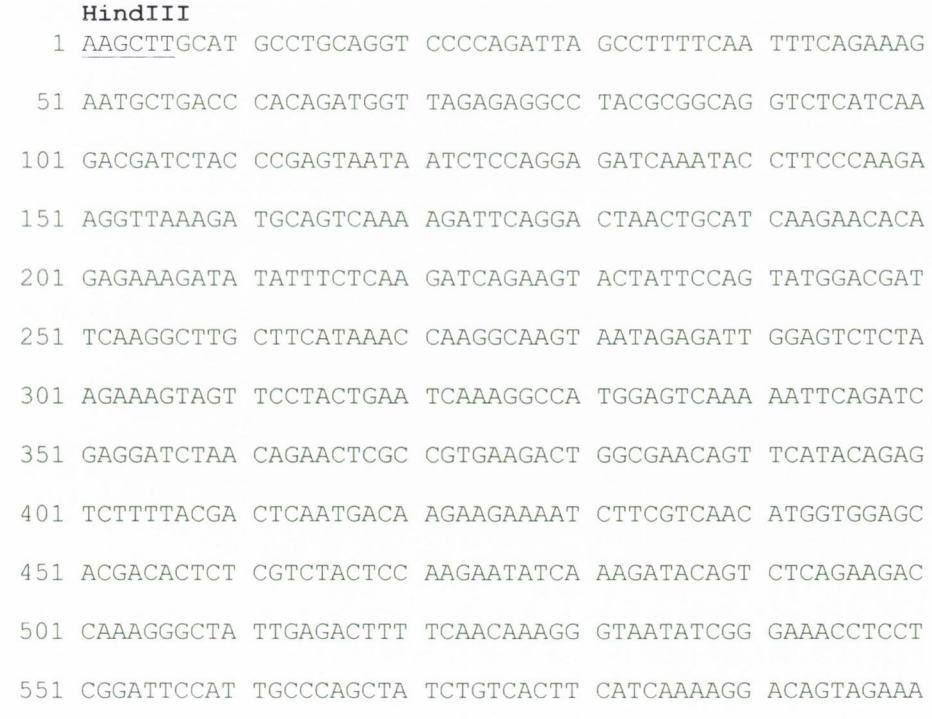
A. PCR product of the atp2-1 beta subunit (**β)** in *Nicotiana plumbaginifolia, β*-60 (lane 2), *β*-90 (lane 3) with flanking *Xba*I */Bam*HI restriction sites, lane 1 is the pUC19 *Sau*3A marker. **B.** PCR product of the chimeric *orfC9* gene sequence with flanking *Bam*HI/*Kpn*I restriction sites (lane 2), lane 1 is the 1 kb marker. **C.** Restriction digestion of pRok219-35S-*β*60-orfC9 with *Bam*HI/*Kpn*I (lane 2), *Xba*I/*Bam*HI (lane 4); restriction digestion of pRok219-*β*90-orfC9 with *Bam*HI/*Kpn*I (lane 3), *Xba*I/*Bam*HI (lane 5); restriction digestion of pRok219-orfC9 with *Bam*HI/*Kpn*I (lane 7). lane 1 is the pUC19 *Sau*3A ladder and lane 6 is the 1 kb ladder. **D.** Restriction digestion with *Hind*III/*Eco*RI of the three binary vector constructs isolated from *Agrobacterium* LBA4404; 1.8 kb pBin19-35S-*β*90-orfC9 (lane 4), 1.7 kb pBin19-35S-*β*60-orfC9 (lane 3), and 1.54 kb pBin19-35S-orfC9 (lane 2). The 1 kb ladder is in lane 1.

A.


HindIII
1 AAGCTTGCAT GCCTGCAGGT CCCCAGATTA GCCTTTTCAA TTTCAGAAAG
51 AATGCTGACC CACAGATGGT TAGAGAGGCC TACGCGGCAG GTCTCATCAA
101 GACGATCTAC CCGAGTAATA ATCTCCAGGA GATCAAATAC CTTCCCAAGA
151 AGGTTAAAGA TGCAGTCAAA AGATTCAGGA CTAACTGCAT CAAGAACACA
201 GAGAAAGATA TATTTCTCAA GATCAGAAGT ACTATTCCAG TATGGACGAT
251 TCAAGGCTTG CTTCATAAAC CAAGGCAAGT AATAGAGATT GGAGTCTCTA
301 AGAAAGTAGT TCCTACTGAA TCAAAGGCCA TGGAGTCAAA AATTCAGATC
351 GAGGATCTAA CAGAACTCGC CGTGAAGACT GGCGAACAGT TCATACAGAG
401 TCTTTTACGA CTCAATGACA AGAAGAAAAT CTTCGTCAAC ATGGTGGAGC
451 ACGACACTCT CGTCTACTCC AAGAATATCA AAGATACAGT CTCAGAAGAC
501 CAAAGGGCTA TTGAGACTTT TCAACAAAGG GTAATATCGG GAAACCTCCT
551 CGGATTCCAT TGCCCAGCTA TCTGTCACTT CATCAAAAGG ACAGTAGAAA
601 AGGAAGGTGG CACCTACAAA TGCCATCATT GCGATAAAGG AAAGGCTATC
651 GTTCAAGATG CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC
701 GAGGAGCATC GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG
CamV35S F
751 TGGATTGATG TGATATCTCC ACTGACGTAA GGGATGACGC ACAATCCCAC
801 TATCCTTCGC AAGACCCTTC CTCTATATAA GGAAGTTCAT TTCATTTGGA
XbaI
851 GAGAACACGG GGGACTCTAG ACCATGGCTT CTCGGAGGCT TCTCGCCTCT
ß90F
901 CTCCTCCGTC AATCGGCTCA ACGTGGCGGC GGTCTAATTT CCCGATCGTT
951 AGGAAACTCC ATCCCTAAAT CCGCTTCACG CGCCTCTTCA CGCGCATCCC
1001 CTAAGGGATT CCTCTTAAAC CGCGCCGTAC AGTACGCTAC CTCCGCAGCG
1051 GCACCGGCAT CTCAGCCATC AACACCACCA AAGTCCGGCA GTGAACCGTC
BamHI
1101 CGGAAAAATT ACCGATGAGT TCACCGGCGC TGGTTCGATC GGGGGATCCA
1151 TGTTAGAAGG AGCGAAACTG GACCCCAAAC CAACGACAAA GAAAATATTC
orfC9F1
1201 CCAGGTTTAT CCGGTTTATT CGGGAGCCTG GGCATTATTA CCTTCGCCGG
orfC9intF
orfC91258
1251 GACAACGGCG GGGATAAAGT ATTTCTTCGG ACTCTCGTCC GTCGAAATGG
1301 CTTTACTAGT GGTTATTTCC ACTTTTTTCA GTGTTGTTTT CCTCTACCTG


1351 ACAACAATGC TCTATAAATA TTCTCGCTTA ATAGAACTCC TTATTTCAAA
orfC9intR
1401 GAACATGGAA CAGGGTGAGA AACAGACTAT CTTTCACGAG TTGGAATTTT
1451 CCCTATATCG TGGGGTCAAA CTCACAGTAA AGGGACGAGA CCAACTCAAG
orfC9R3(R2.1) KpnI
1501 GGAGAAACAA CGCATGAATA AGGTACCGAG CTCGAATTTC CCCGATCGTT
1551 CAAACATTTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGCCGGTCTT
1601 GCGATGATTA TCATATAATT TCTGTTGAAT TACGTTAAGC ATGTAATAAT
1651 TAACATGTAA TGCATGACGT TATTTATGAG ATGGGTTTTT ATGATTAGAG
1701 TCCCGCAATT ATACATTTAA TACGCGATAG AAAACAAAAT ATAGCGCGCA
1751 AACTAGGATA AATTATCGCG CGCGGTGTCA TCTATGTTAC TATGTTACTA
EcoRI
1801 GATCGGGAAT TC

B.


HindIII
1 AAGCTTGCAT GCCTGCAGGT CCCCAGATTA GCCTTTTCAA TTTCAGAAAG
51 AATGCTGACC CACAGATGGT TAGAGAGGCC TACGCGGCAG GTCTCATCAA
101 GACGATCTAC CCGAGTAATA ATCTCCAGGA GATCAAATAC CTTCCCAAGA
151 AGGTTAAAGA TGCAGTCAAA AGATTCAGGA CTAACTGCAT CAAGAACACA
201 GAGAAAGATA TATTTCTCAA GATCAGAAGT ACTATTCCAG TATGGACGAT
251 TCAAGGCTTG CTTCATAAAC CAAGGCAAGT AATAGAGATT GGAGTCTCTA
301 AGAAAGTAGT TCCTACTGAA TCAAAGGCCA TGGAGTCAAA AATTCAGATC
351 GAGGATCTAA CAGAACTCGC CGTGAAGACT GGCGAACAGT TCATACAGAG
401 TCTTTTACGA CTCAATGACA AGAAGAAAAT CTTCGTCAAC ATGGTGGAGC
451 ACGACACTCT CGTCTACTCC AAGAATATCA AAGATACAGT CTCAGAAGAC
501 CAAAGGGCTA TTGAGACTTT TCAACAAAGG GTAATATCGG GAAACCTCCT
551 CGGATTCCAT TGCCCAGCTA TCTGTCACTT CATCAAAAGG ACAGTAGAAA


601 AGGAAGGTGG CACCTACAAA TGCCATCATT GCGATAAAGG AAAGGCTATC
651 GTTCAAGATG CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC
701 GAGGAGCATC GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG
CamV35S F
751 TGGATTGATG TGATATCTCC ACTGACGTAA GGGATGACGC ACAATCCCAC
801 TATCCTTCGC AAGACCCTTC CTCTATATAA GGAAGTTCAT TTCATTTGGA
XbaI
851 GAGAACACGG GGGACTCTAG ACCATGGCTT CTCGGAGGCT TCTCGCCTCT
ß60F
901 CTCCTCCGTC AATCGGCTCA ACGTGGCGGC GGTCTAATTT CCCGATCGTT
951 AGGAAACTCC ATCCCTAAAT CCGCTTCACG CGCCTCTTCA CGCGCATCCC
1001 CTAAGGGATT CCTCTTAAAC CGCGCCGTAC AGTACGCTAC CTCCGCAGCG
BamHI
1051 GCAGGATCCA TGTTAGAAGG AGCGAAACTG GACCCCAAAC CAACGACAAA
orfC9F.1
1101 GAAAATATTC CCAGGTTTAT CCGGTTTATT CGGGAGCCTG GGCATTATTA
orfC9intF
orfC9R1258
1151 CCTTCGCCGG GACAACGGCG GGGATAAAGT ATTTCTTCGG ACTCTCGTCC
1201 GTCGAAATGG CTTTACTAGT GGTTATTTCC ACTTTTTTCA GTGTTGTTTT
1251 CCTCTACCTG ACAACAATGC TCTATAAATA TTCTCGCTTA ATAGAACTCC
1301 TTATTTCAAA GAACATGGAA CAGGGTGAGA AACAGACTAT CTTTCACGAG
orfC9intR
1351 TTGGAATTTT CCCTATATCG TGGGGTCAAA CTCACAGTAA AGGGACGAGA
orfC9(R3)R2.1 KpnI
1401 CCAACTCAAG GGAGAAACAA CGCATGAATA AGGTACCGAG CTCGAATTTC
1451 CCCGATCGTT CAAACATTTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT
1501 TGCCGGTCTT GCGATGATTA TCATATAATT TCTGTTGAAT TACGTTAAGC
1551 ATGTAATAAT TAACATGTAA TGCATGACGT TATTTATGAG ATGGGTTTTT
1601 ATGATTAGAG TCCCGCAATT ATACATTTAA TACGCGATAG AAAACAAAAT
1651 ATAGCGCGCA AACTAGGATA AATTATCGCG CGCGGTGTCA TCTATGTTAC
EcoRI
1701 TATGTTACTA GATCGGGAAT TC

C.

HindIII

```
   1 AAGCTTGCAT GCCTGCAGGT CCCCAGATTA GCCTTTTCAA TTTCAGAAAG
  51 AATGCTGACC CACAGATGGT TAGAGAGGCC TACGCGGCAG GTCTCATCAA
 101 GACGATCTAC CCGAGTAATA ATCTCCAGGA GATCAAATAC CTTCCCAAGA
 151 AGGTTAAAGA TGCAGTCAAA AGATTCAGGA CTAACTGCAT CAAGAACACA
 201 GAGAAAGATA TATTTCTCAA GATCAGAAGT ACTATTCCAG TATGGACGAT
 251 TCAAGGCTTG CTTCATAAAC CAAGGCAAGT AATAGAGATT GGAGTCTCTA
 301 AGAAAGTAGT TCCTACTGAA TCAAAGGCCA TGGAGTCAAA AATTCAGATC
 351 GAGGATCTAA CAGAACTCGC CGTGAAGACT GGCGAACAGT TCATACAGAG
 401 TCTTTTACGA CTCAATGACA AGAAGAAAAT CTTCGTCAAC ATGGTGGAGC
 451 ACGACACTCT CGTCTACTCC AAGAATATCA AAGATACAGT CTCAGAAGAC
 501 CAAAGGGCTA TTGAGACTTT TCAACAAAGG GTAATATCGG GAAACCTCCT
 551 CGGATTCCAT TGCCCAGCTA TCTGTCACTT CATCAAAAGG ACAGTAGAAA
 601 AGGAAGGTGG CACCTACAAA TGCCATCATT GCGATAAAGG AAAGGCTATC
 651 GTTCAAGATG CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC
 701 GAGGAGCATC GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG
     CamV35S F ------>
 751 TGGATTGATG TGATATCTCC ACTGACGTAA GGGATGACGC ACAATCCCAC
 801 TATCCTTCGC AAGACCCTTC CTCTATATAA GGAAGTTCAT TTCATTTGGA
                    XbaI  BamHI      orfC9F.1 ------>
 851 GAGAACACGG GGGACTCTAG AGGATCCATG TTAGAAGGAG CGAAACTGGA
                                     orfC9Fint ------>
 901 CCCCAAACCA ACGACAAAGA AAATATTCCC AGGTTTATCC GGTTTATTCG
     <------ orfC91258
 951 GGAGCCTGGG CATTATTACC TTCGCCGGGA CAACGGCGGG GATAAAGTAT
1001 TTCTTCGGAC TCTCGTCCGT CGAAATGGCT TTACTAGTGG TTATTTCCAC
1051 TTTTTTCAGT GTTGTTTTCC TCTACCTGAC AACAATGCTC TATAAATATT
1101 CTCGCTTAAT AGAACTCCTT ATTTCAAAGA ACATGGAACA GGGTGAGAAA
                    <------ orfC9intR
1151 CAGACTATCT TTCACGAGTT GGAATTTTCC CTATATCGTG GGGTCAAACT
                            <------ orfC9(R3) R2.1
1201 CACAGTAAAG GGACGAGACC AACTCAAGGG AGAAACAACG CATGAATAAG
     KpnI
1251 GTACCGAGCT CGAATTTCCC CGATCGTTCA AACATTTGGC AATAAAGTTT
1301 CTTAAGATTG AATCCTGTTG CCGGTCTTGC GATGATTATC ATATAATTTC
```

```
1351 TGTTGAATTA CGTTAAGCAT GTAATAATTA ACATGTAATG CATGACGTTA

1401 TTTATGAGAT GGGTTTTTAT GATTAGAGTC CCGCAATTAT ACATTTAATA

1451 CGCGATAGAA AACAAAATAT AGCGCGCAAA CTAGGATAAA TTATCGCGCG
                                              EcoRI
1501 CGGTGTCATC TATGTTACTA TGTTACTAGA TCGGGAATTC
```

Figure 5.3 Confirmation of the sequence within the *Hind*III/*Eco*RI restriction sites for the three constructs in pBin19

A. Is the pBin19-35S-*ß*90-orfC9 constuct. **B.** Is the pBin19-35S-*ß*60-orfC9 construct. **C.** Is the pBin19-35S-orfC9 construct. The green colour type indicates the sequence for CaMV35S promoter, the blue colour type is for the trans-peptide (*ß*-90 or *ß*-60) sequence, the red colour type is for the orfC9 sequence, and the black color type is for the NosT terminator sequence. The restriction sites for *Bam*HI, *Kpn*I, *Xba*I are underlined, and the primers used for sequencing purposes are highlighted with arrows.

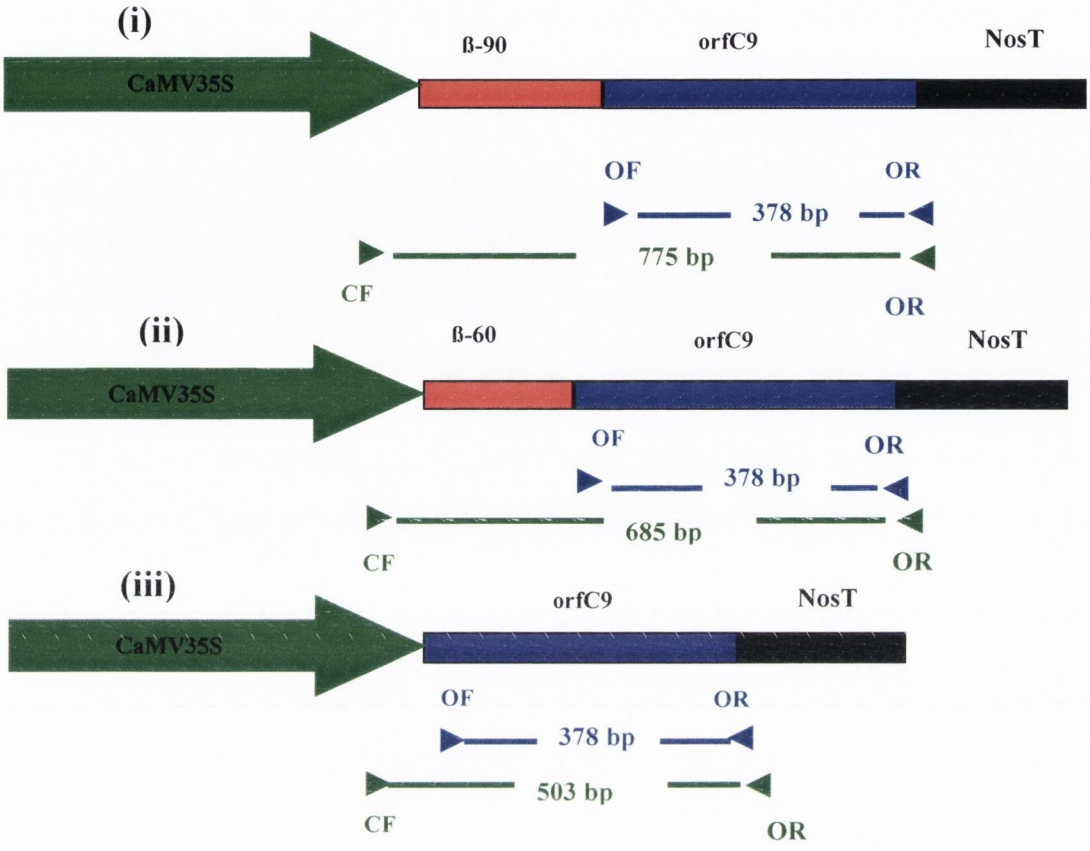


Figure 5.4 PCR analysis confirming the insert for the three different transgenes. The three different transgene constructs (**i, ii, iii**) with the CaMV35S promoter , *ß*-90 or *ß*-60 pre-sequence where applicable, the orfC9 gene and the NosT terminator are highlighted. PCR product using the CaMV35SF(**CF**) and orfC9R(**OR**) primer set amplified fragment sizes of 775 bp, 685 bp, 503 bp in the three constructs represented by (**i, ii, iii**), respectively. The orfC9 fragment of 378 bp in size is amplified in the three different transgenes, and the primers used for their amplification is the orfC9F(**OF**) and orfC9R(**OR**) primer set.

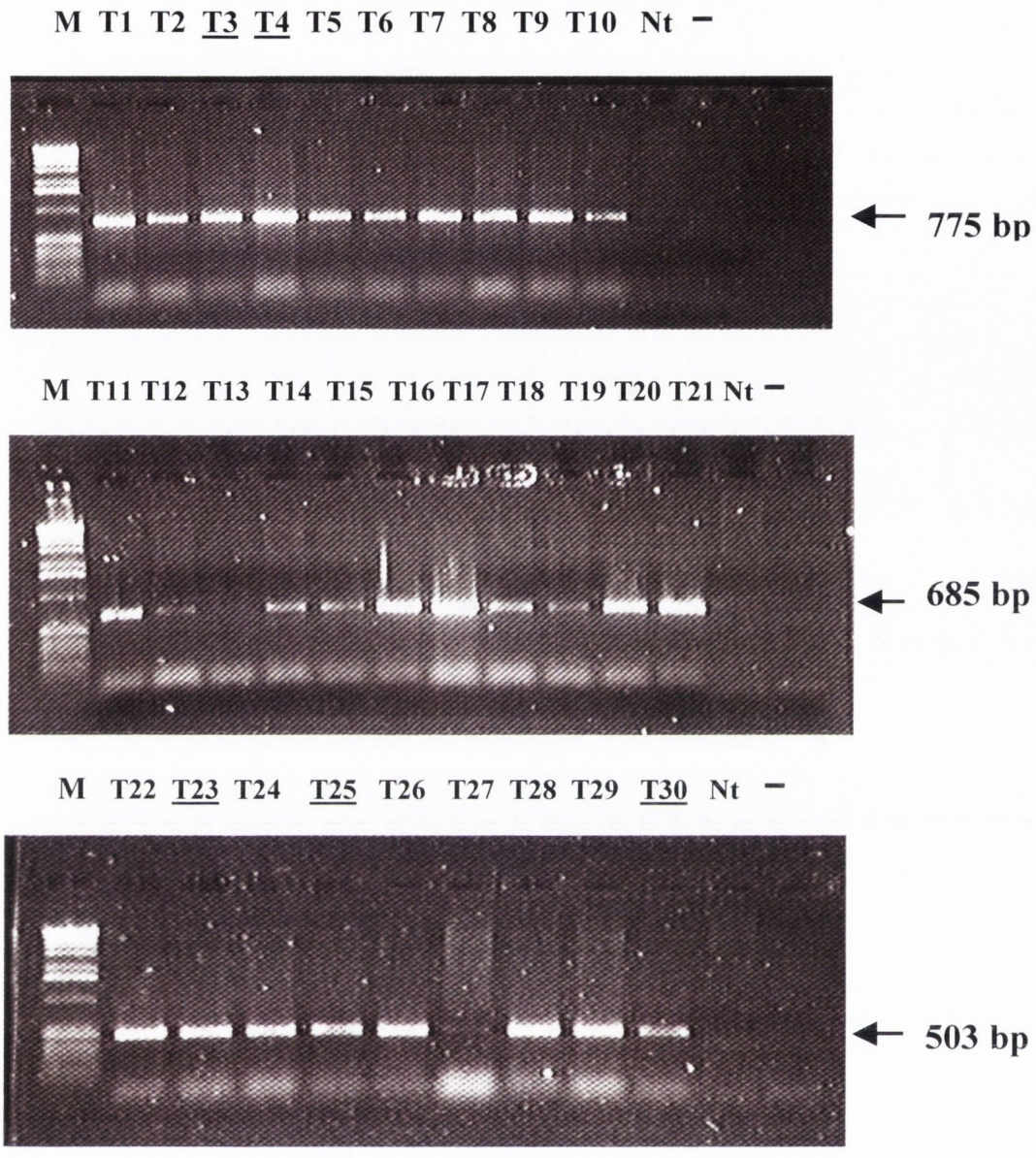


Figure 5.5 PCR confirmation of the 35S-β90-orfC9, 35S-β60-orfC9, 35S-orfC9 region in nuclear genome of *N. tabacum.*

PCR analysis using the primer set CaMV35SF/orfC9R for the transformed tobacco plants with constructs 35S-*β*90-orfC9 (T1 to T10), 35S-*β*60-orfC9 (T11 to T21), 35S-orfC9 (T22 to T30), the non-transformed tobacco plants (Nt), and the negative control (−). Lane M refers to the 1 kb marker.

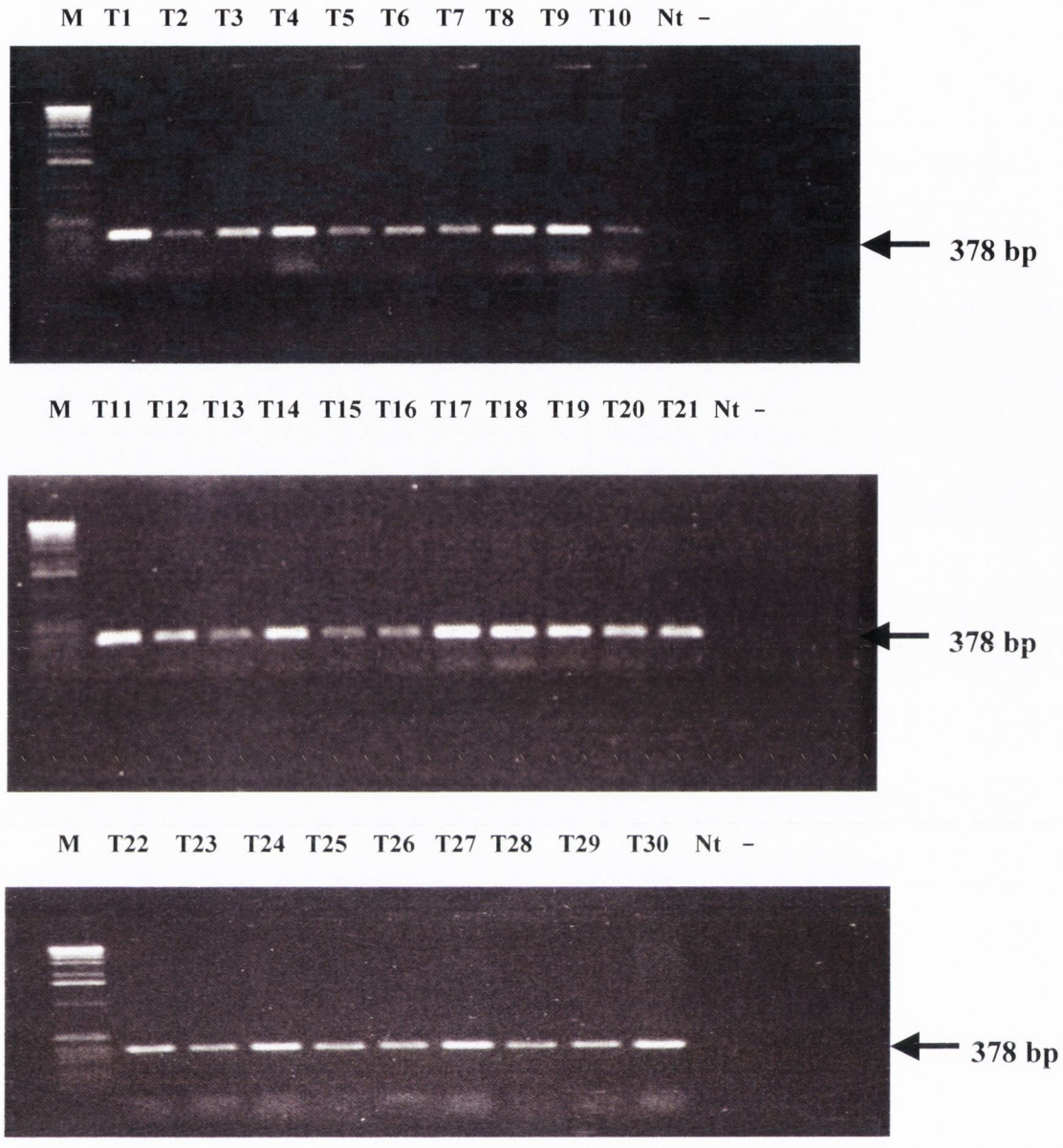


Figure 5.6 PCR confirmation of the 35S-β90-orfC9, 35S-β60-orfC9, 35S-orfC9 region in nuclear genome of *N. tabacum.*

PCR analysis using the primer set orfC9F/orfC9R for the transformants 35S-*β*90-orfC9 (T1 to T10), 35S-*β*60-orfC9 (T11 to T21), 35S-orfC9 (T22 to T30), the non-transformed tobacco plants (Nt), and the negative control (−). Lane M refers to the 1 kb marker

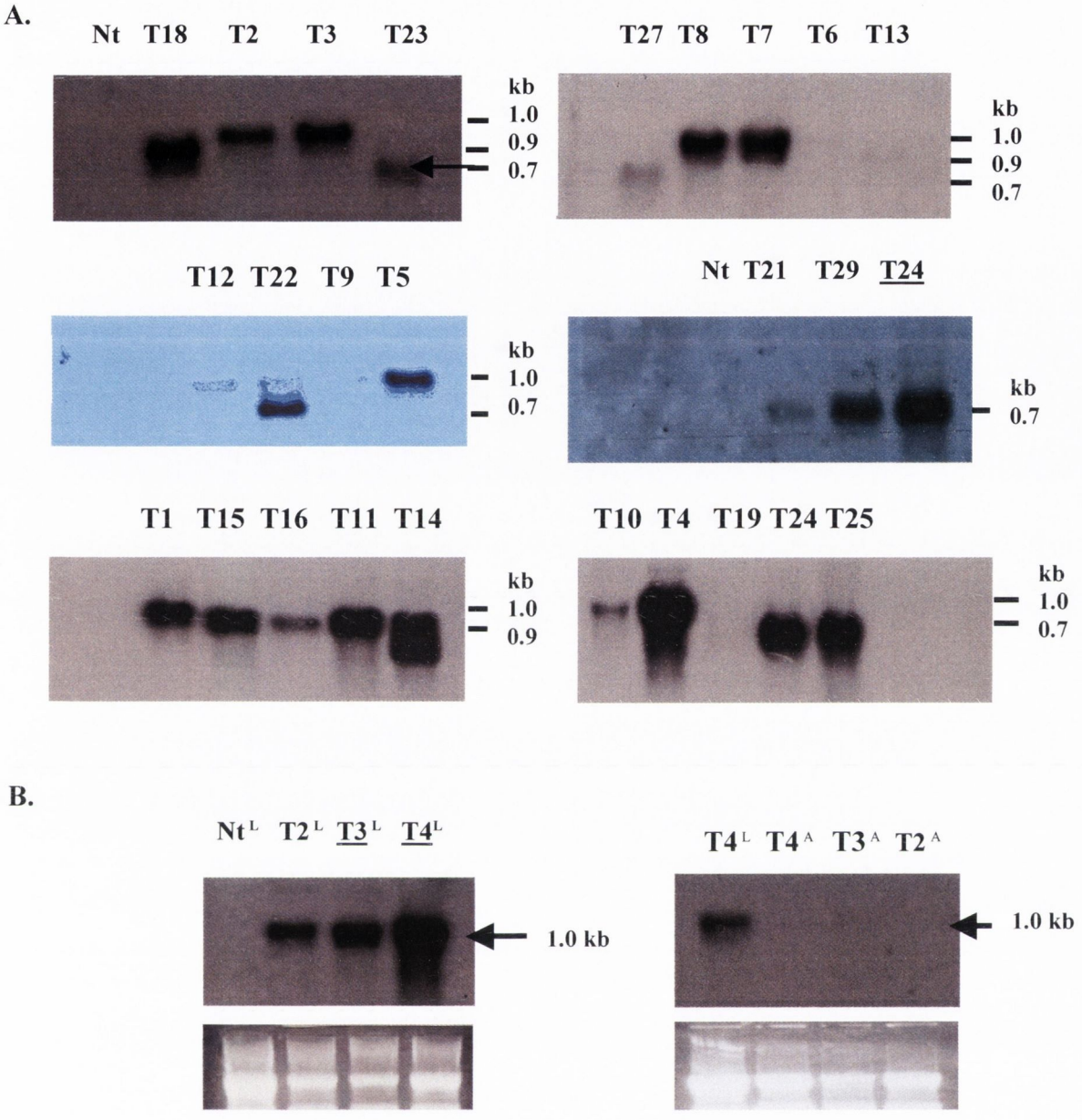


Figure 5.7 Northern blots of To transformed lines probed with orfC9

A. Total RNA from transformed tobacco To leaf lines of CaMV35S-β90-orfC9 (T1 to T10) 1.0 kb, CaMV35S-β60-orfC9 (T13, T18,T11, T14, T15, T16) 0.9 kb, and CaMV35S-orfC9 (T22, T23, T24, T25, T26, T28, T29, T30) 0.7 kb probed with orfC9. Nt is total RNA from non-transformed tobacco. **B.** Total RNA from leaf tissue for tobacco lines (Nt, T2, T3, T4)^{L} and total RNA from anther tissue for tobacco lines (T2, T3 and T4)^{A} probed with orfC9 giving a transcript size of 1.0 kb in leaf tissue only.

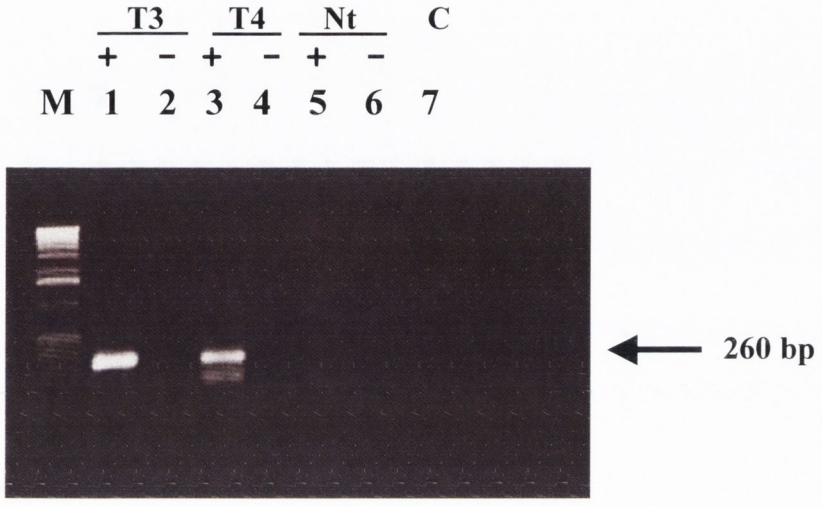


Figure 5.8 PCR amplication from cDNA anther tissue in linesT3 and T4

A PCR reaction showing the amplification of a PCR product of 260 bp in size, with the internal orfC9 primer set (orfc9intF and orfC9intR) from the cDNA template for anther tissue from the **To** transformed lines T3 and T4. Lane 1, lane 3 and lane 5 refer to reverse transcribed (+) cDNA lines of the transformed T3 and T4 lines and of the non-transformed (Nt) line. Lanes 2, 4 and 6 refers to the non-reverse transcribed (-) cDNA, and lane 7 is negative the control, namely minus DNA template (**C**). Lane M is the 1 kb marker.

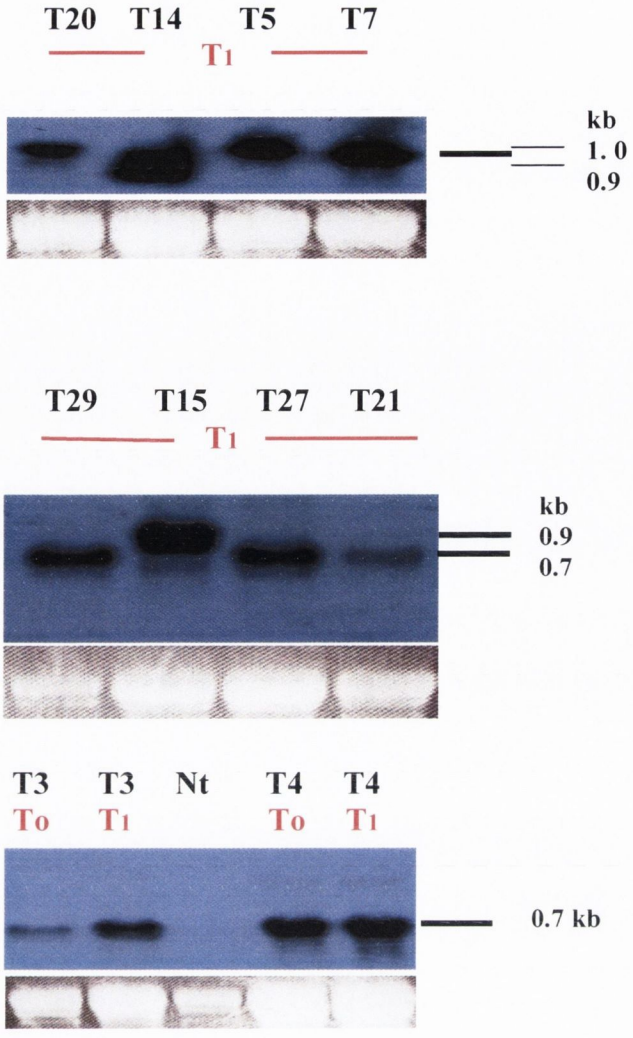


Figure 5.9 Northern blots of T1 transformed lines probed with orfC9

Transformed lines of T1 is represented by T3, T4, T5 and T7 for the CaMV35S-β90-orfC9 constructs; T14, T15, T20 for the CaMV35S-β60-orfC9 construct; T21, T27, T29 for the CaMV35S-orfC9 construct. The To transformed lines T3, T4 the CaMV35S-β90-orfC9 transformants are used as positive controls whereas the non-transformed (Nt) line is the negative control.

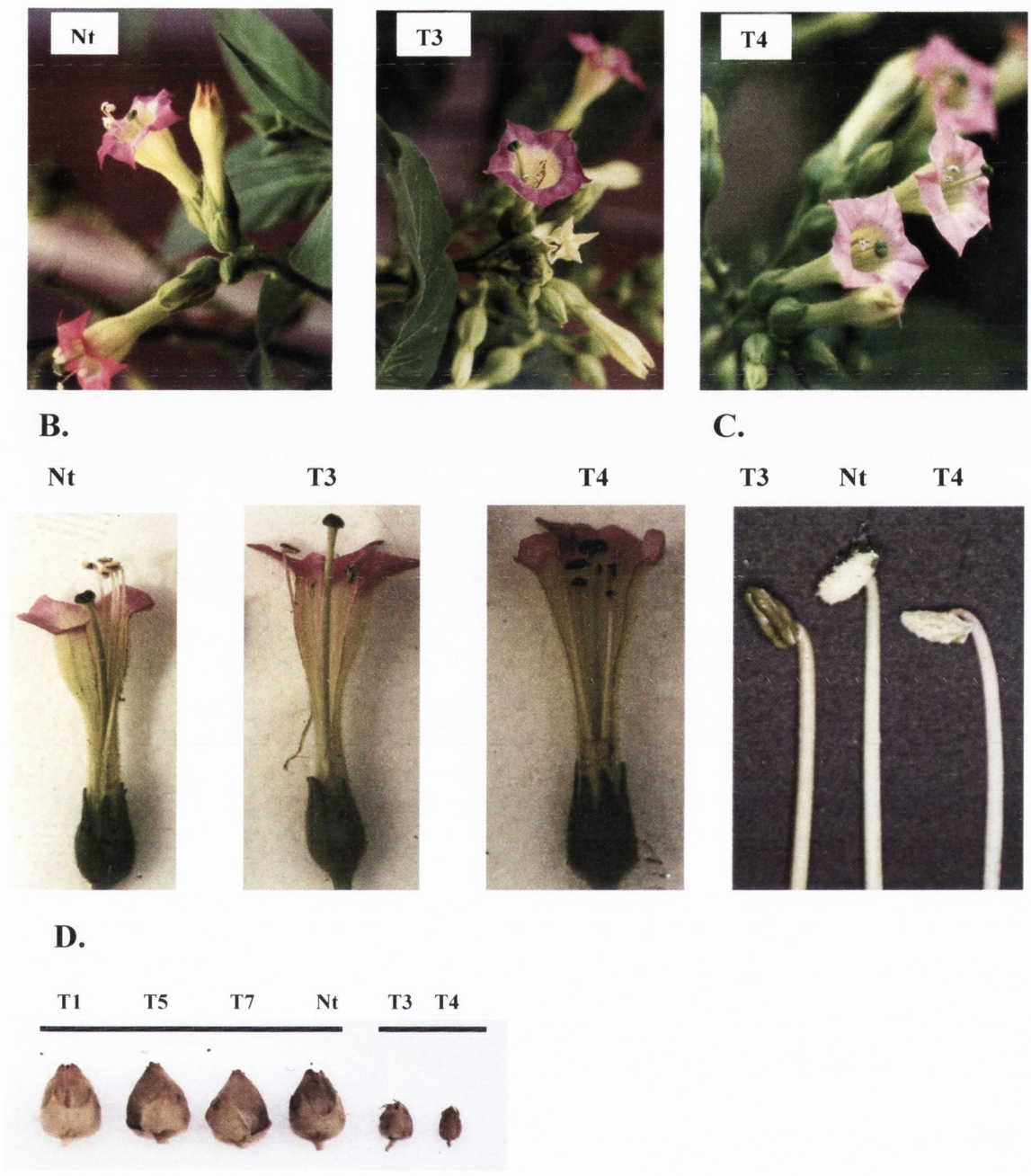


Figure 5.10 Flower Morphology and seed set for the To generation.

A. The flower morphology of the lines T3 and T4 shows that the anthers are recessed below the stigma compared to non-transformed flower-Nt. **B.** The lines T3 and T4 illustrates a sterile phenotype with the anthers recessed beneath the stigma in comparison to the fertile non-transformed line Nt. **C.** Shows the stamens for the lines T3, T4 and Nt. The pollen grains on the dehisced anthers for T4 are significantly reduced in number relative to the non-transformed line Nt, with non-dehisced anthers in T3 line showing no visible pollen grains. **D.** The reduced fruit capsule size for the lines T3 and T4 in comparison to fruit capsules from other fertile transformed lines T1, T5, T7 and the fertile non-transformed line Nt.

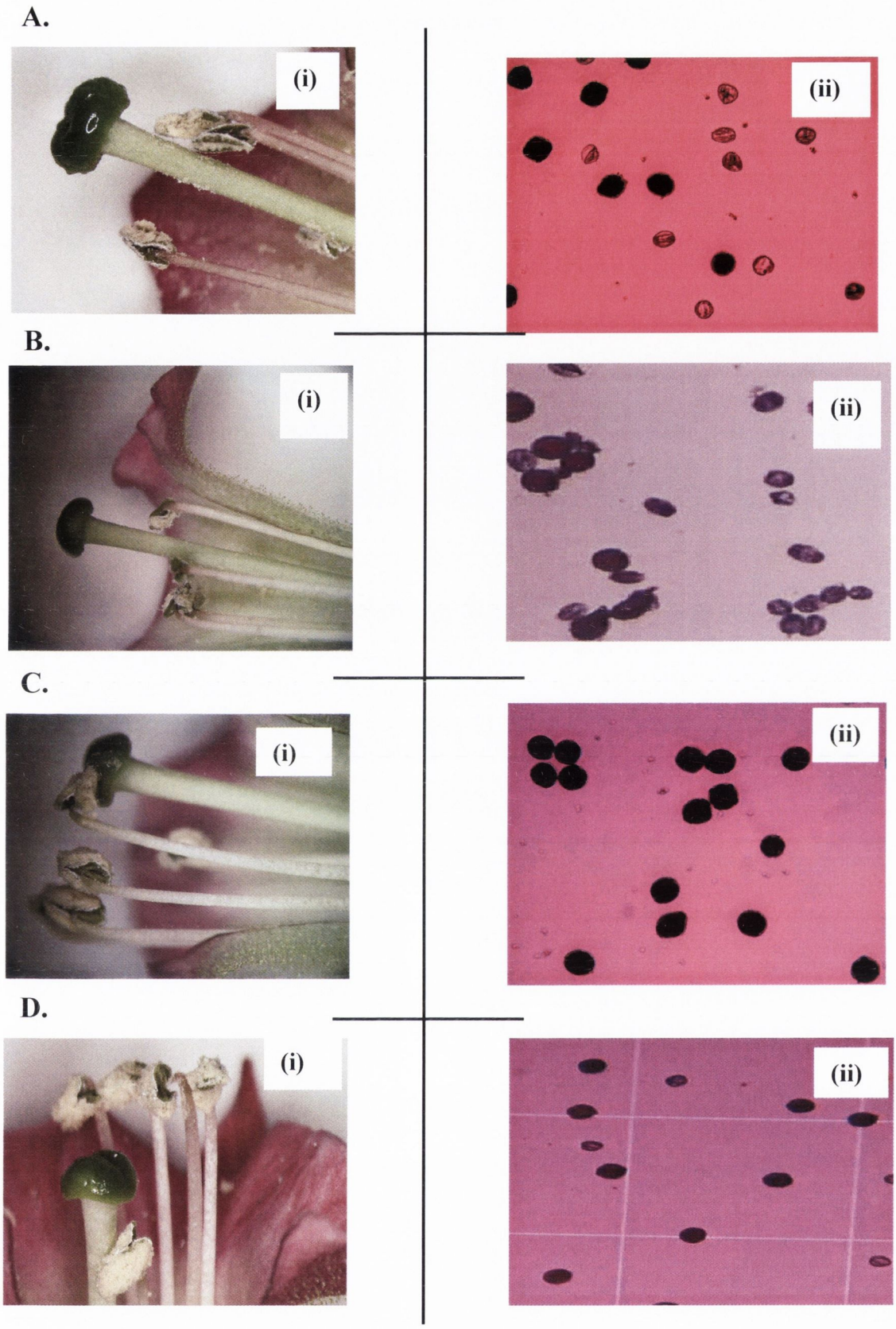


Figure 5.11 Characterisation of the male sterile phenotype in the To generation.

A and B show flowers from the transgenic lines T3 and T4 having a semi-sterile phenotype with (i) anthers recessed well below the stigma and in (ii) a mixture of pollen grains stained with Potassium Iodide illustrating semi-sterile pollen grains (empty pollen sacs). **C** is the non-transformed line Nt with (i) non-recessed anthers (ii) full pollen sacs, fully stained with Potassium Iodide. **D** is -fertile transformed line with (i) non-recessed anthers (ii) full pollen sacs, fully stained with Potassium Iodide (KI).

A.



T3 (T_1) x Nt

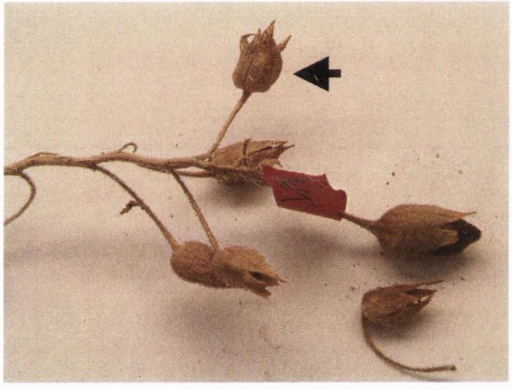


T4 (T_1) x Nt

B.

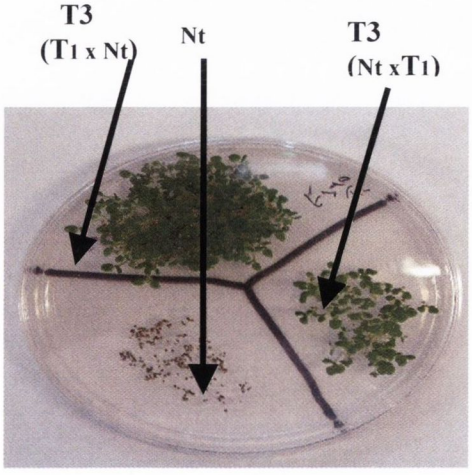


Nt x T3(T_1)



Nt x T4 (T_1)

C.



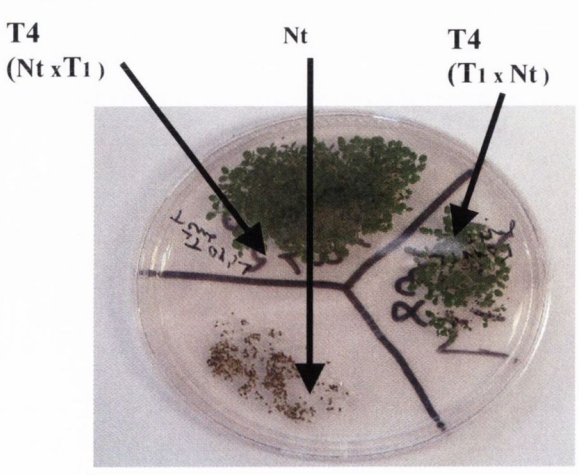


Figure 5.12 Production of fruit capsules by crossing semi-sterile phenotype flowers with non-transformed line Nt to verify the semi-sterile phenotype.

A. Examination of male fertility in the lines T3 and T4; by separately crossing pollen from the lines T3 and T4 onto the stigma of the non-transformed line Nt. The self-fertilised fruit capsule is highlighted by the symbol ► . **B.** Examination of female fertility in the lines T3 and T4 by crossing the stigma with fertile pollen from the non-transformed line Nt. The self-fertilised fruit capsule is highlighted by the symbol ► . **C.** Seeds from the cross lines T3 (**T1 x Nt**) , T3 (**Nt xT1**), T4 (**T1 x Nt**) and T4 (**Nt xT1**) generation were plated onto MS kanamycin medium (500 μg/ml). **Nt** is the negative control and refers to seeds from a non-transformed *N. tabacum* (var. Samsun NN cultivar) plant.

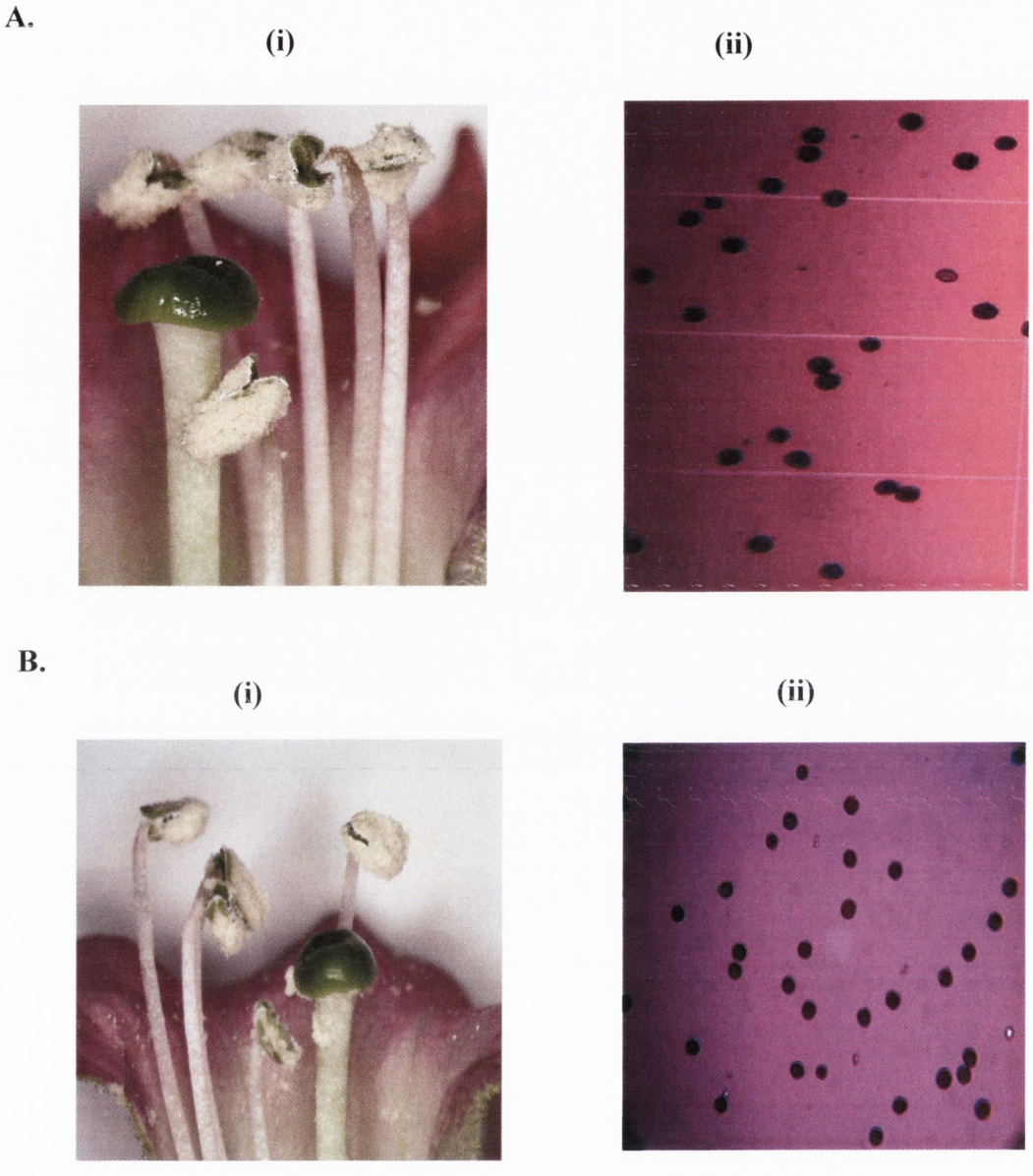


Figure 5.13 Characterisation of the male sterile phenotype in the To generation of the 35S-β60-orfC9 and 35SS-orfC9 transformants.

A. Flower morphology of the 35S-*β*60-orfC9 transformant with (i) anthers above the stigma and (ii) pollen grains stained with Potassium Iodide (Lugol solution). **B.** Flower morphology of the 35S-orfC9 transformant with (i) anthers above the stigma and (ii) pollen grains stained with Potassium Iodide (Lugol solution).

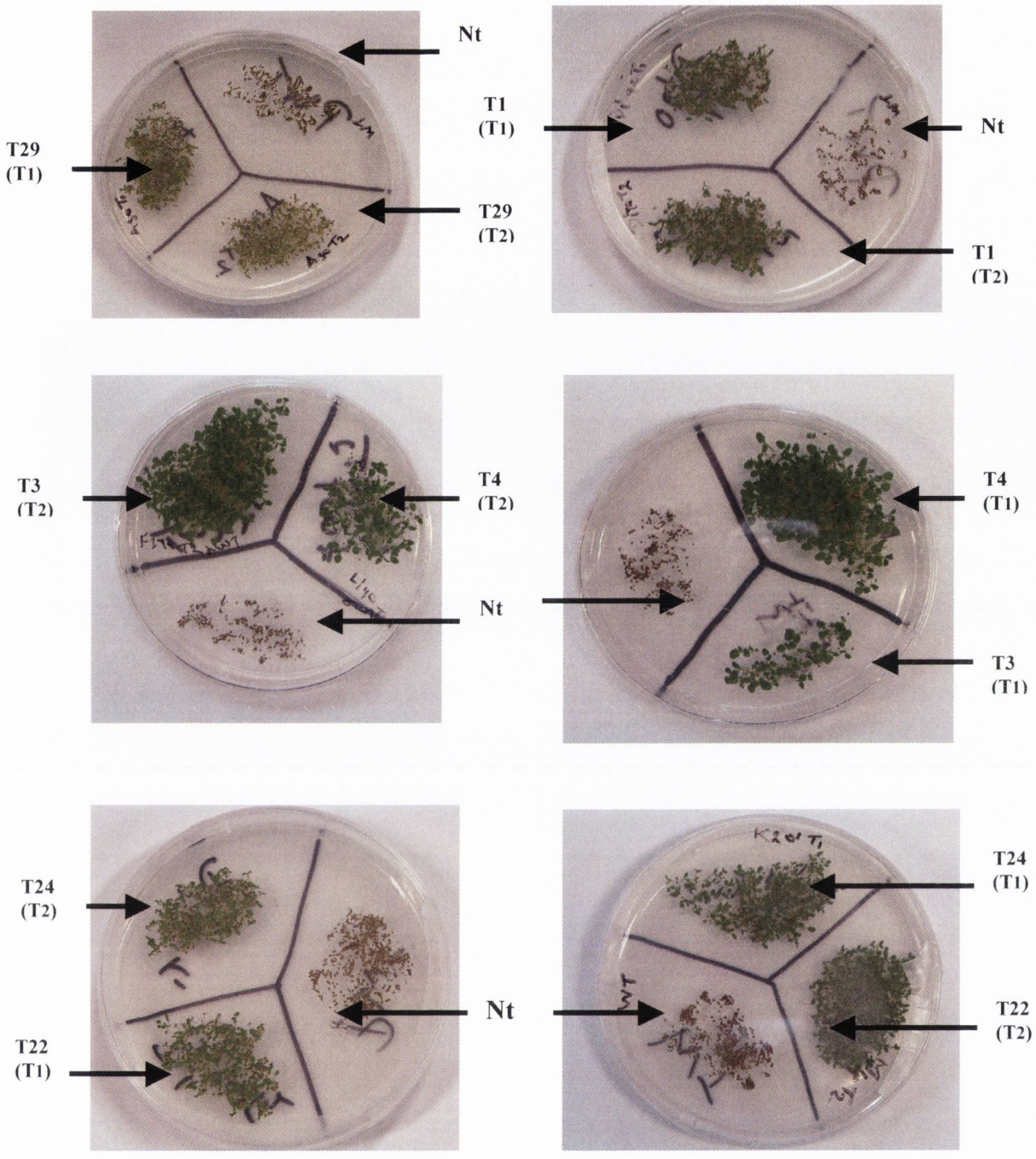
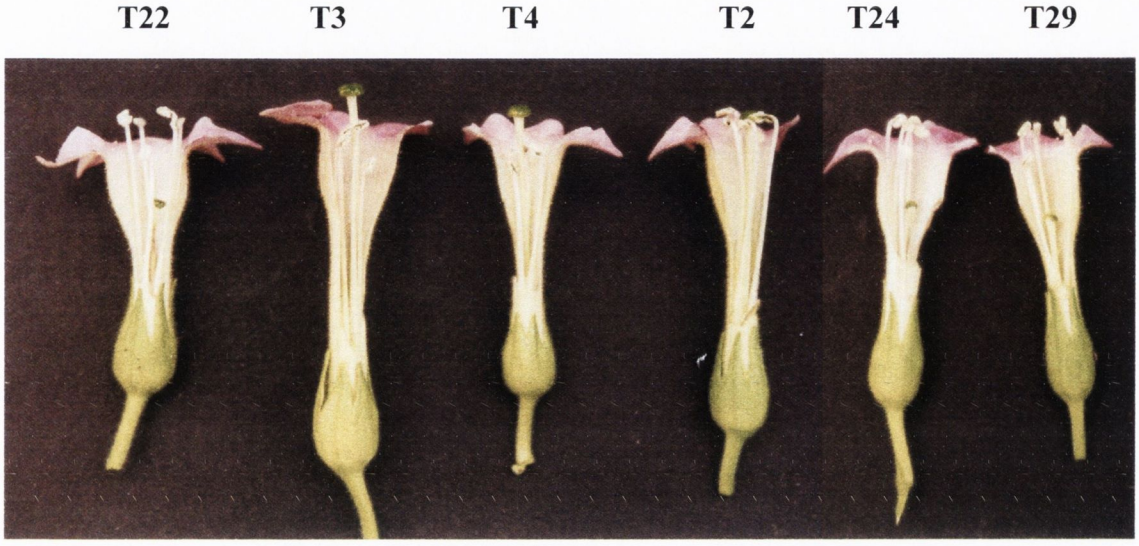


Figure 5.14 Germination of Seedlings in the T1 and T2 generation for the 35S-β90-orfC9 and 35S-orfC9 transformants.

The germination of seeds on MS medium containing Kanamycin (500 μg/ml) for the T1 and T2 genera in the lines T1, T3, T4, T22, T24, T29. Seeds from the non-transformed line (Nt) were unable to germinate.

A.



B.

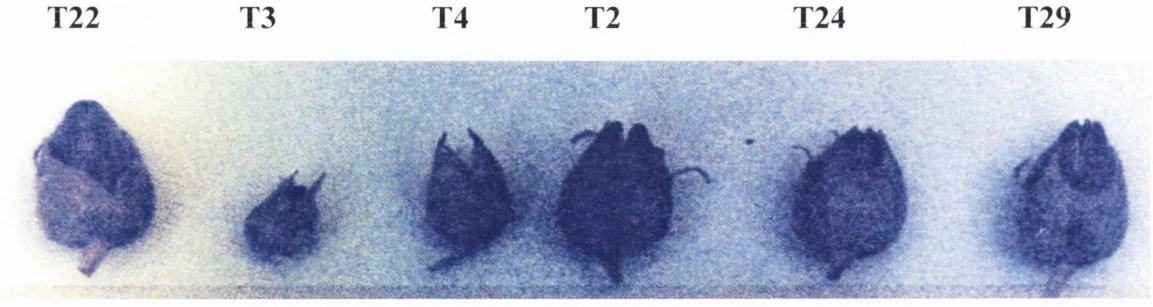
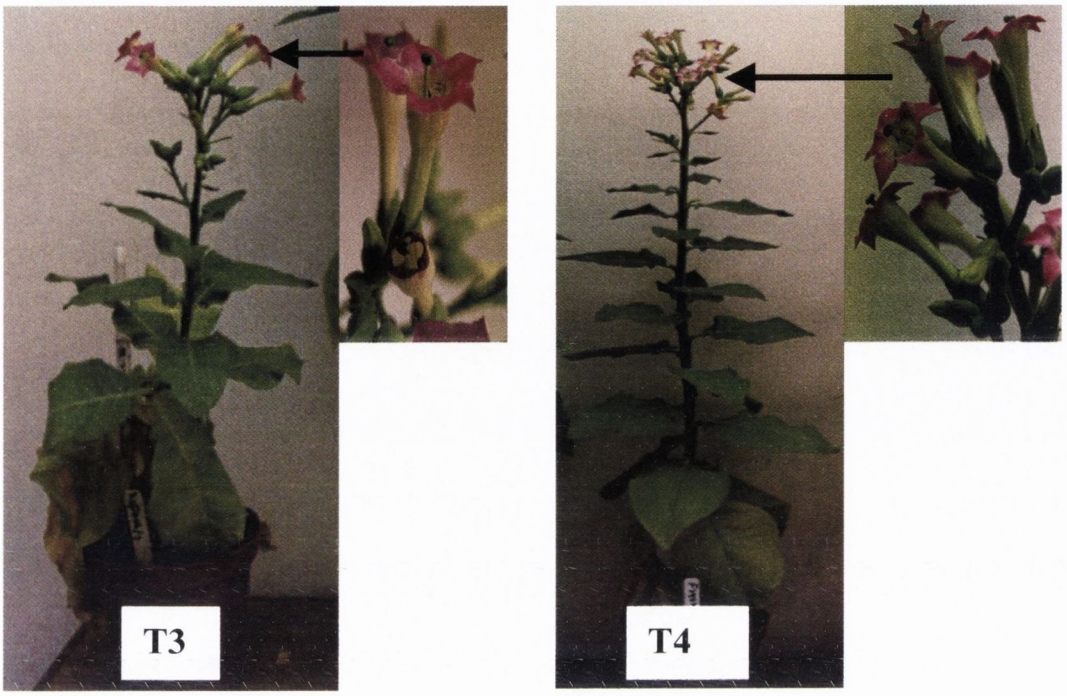


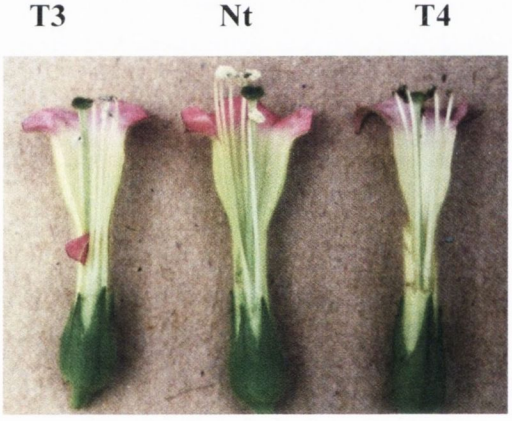
Figure 5.15 Examination of the phenotype of the T1 generation

A. Illustrates the morphology of flowers for the transformed lines T1, T3, T4, T22, T24, T29.**B**. Fruit capsules of Transgenic lines T2, T22, T24, T29 with seeds and the lines T3, T4, with reduced seeds after self fertilisation.

A.



B.



C.



Figure 5.16 Transformed plants from the T_2 genera were grown to flowering.

A. Transgenic plants from the lines T3 and T4 for the T2 genera were allowed to develop into flowering plants. **B.** The anthers illustrated a male-sterile phenotype for the lines T3 and T4 when compared to anthers from a fertile non-transformed line (Nt) plants. **C.** Fruit capsules showed a reduced seed set for those that resulted from self-fertilisation whereas those fertilised with fertile pollen from the Nt showed a normal seed set highlighted by the arrow symbol ◄ (taped red).

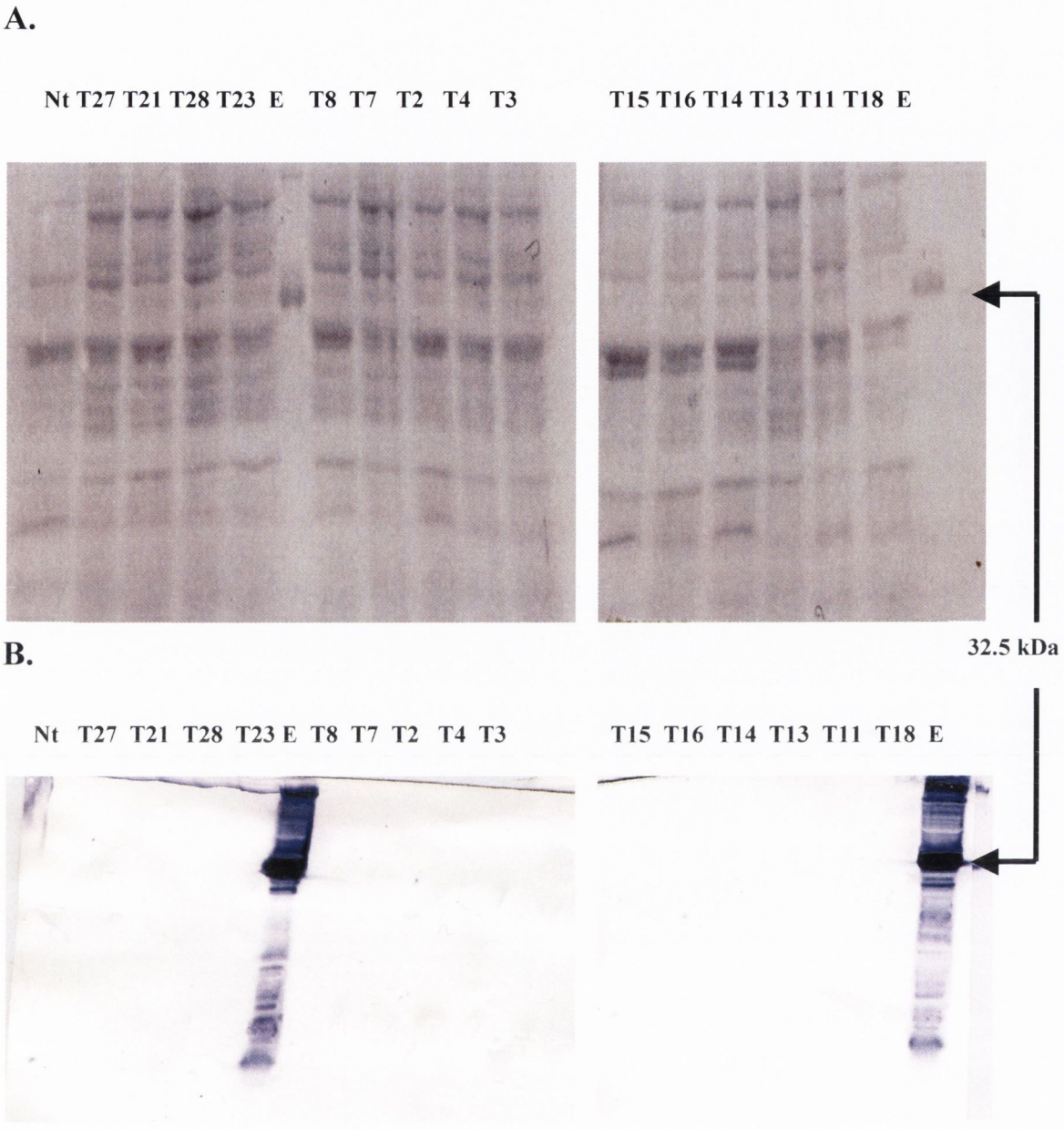


Figure 5.17 Western blot of the transgenic lines from *N. Tabacum* hybridised with the anti-orfC9.

A. Coomassie stained 14 % SDS-PAGE of total protein of approx. 50 µg from leaf tissue for the non-transformed line (Nt), 35S-β90-orfC9 transformed lines T2, T3 T4,T7, T8, the 35S-β60-orfC9 transformed lines T11, T13. T14, T15, T16, T18, and the 35S-orfC9 transformed lines T21, T23, T27, T28. The insoluble protein fraction (Ip) for the *E.coli* pet32aorfC9 fraction is represented by lane E. **B.** Western blot of the coomassie stained gel in 5.17A. The. electroblot is probed with TrxHisSorfC9 antibody (Ab) at a dilution of 1/1000.

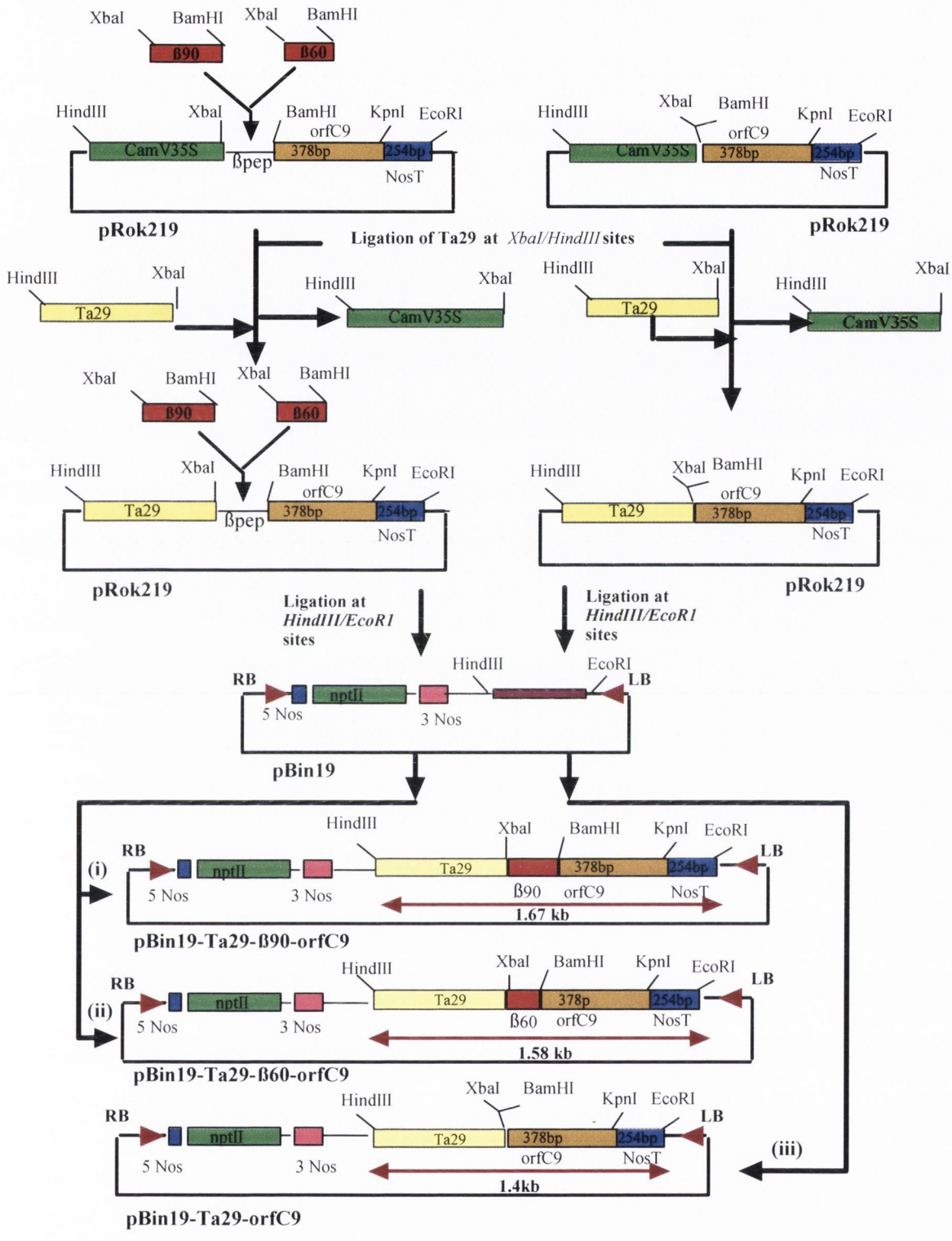


Figure 5.18 Construction of the Ta29 pollen promoter , the chimeric orfC9 gene and the transit peptide genes ß90, ß60 in the binary vector pBin19.

(i) pBin19-Ta29-*ß*90-orfC9 (ii) pBin19-Ta29-*ß*60-orfC9 (iii) pBin19-Ta29-orfC9, (diagram not to scale). The Ta29 promoter, the ß-peptide sequences encoding the *ß*-90 and *ß*-60 NH_2-terminal amino acid residues of the ATPase ß-precursor, and the CMS accociated gene orfC9.

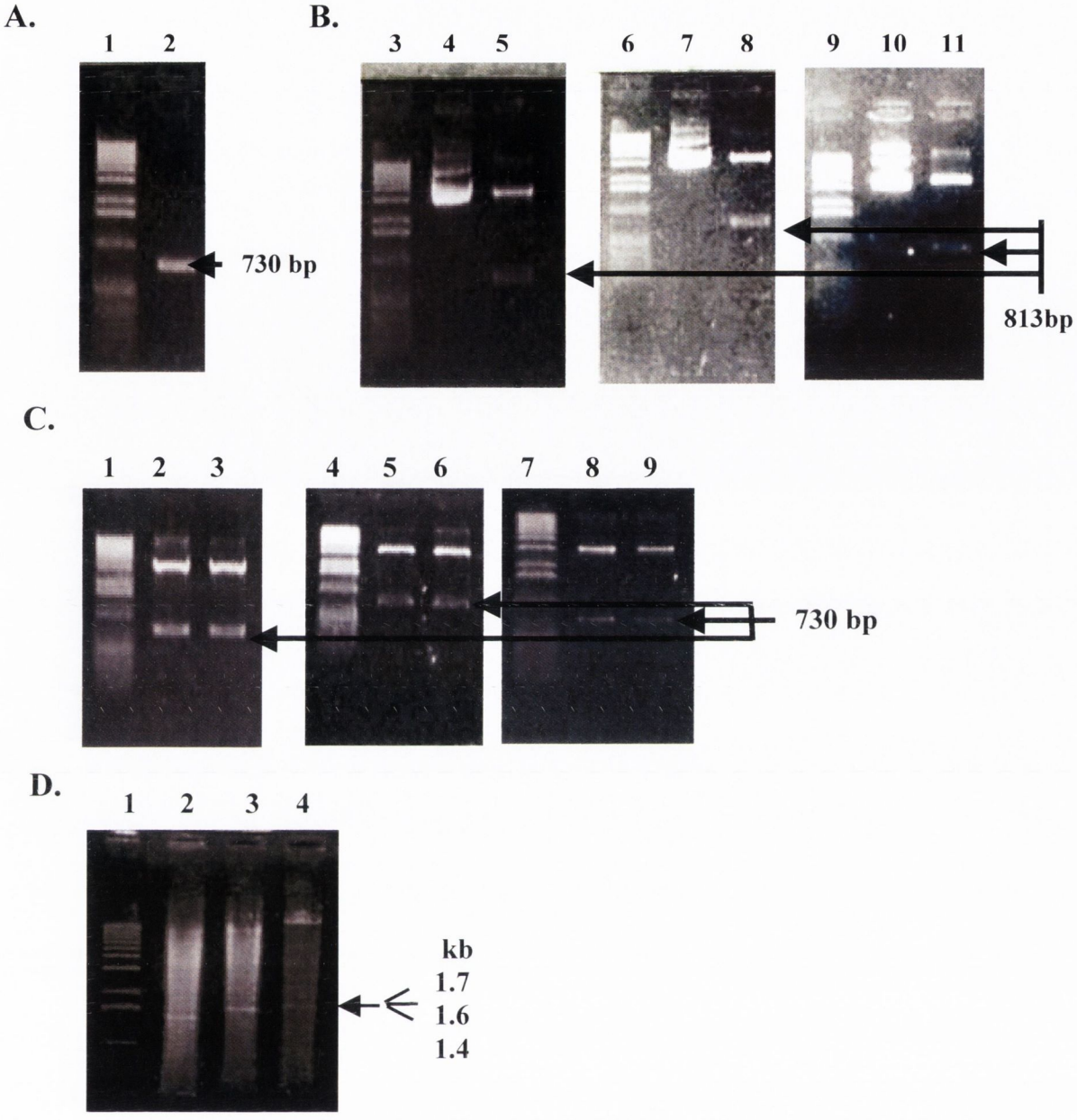
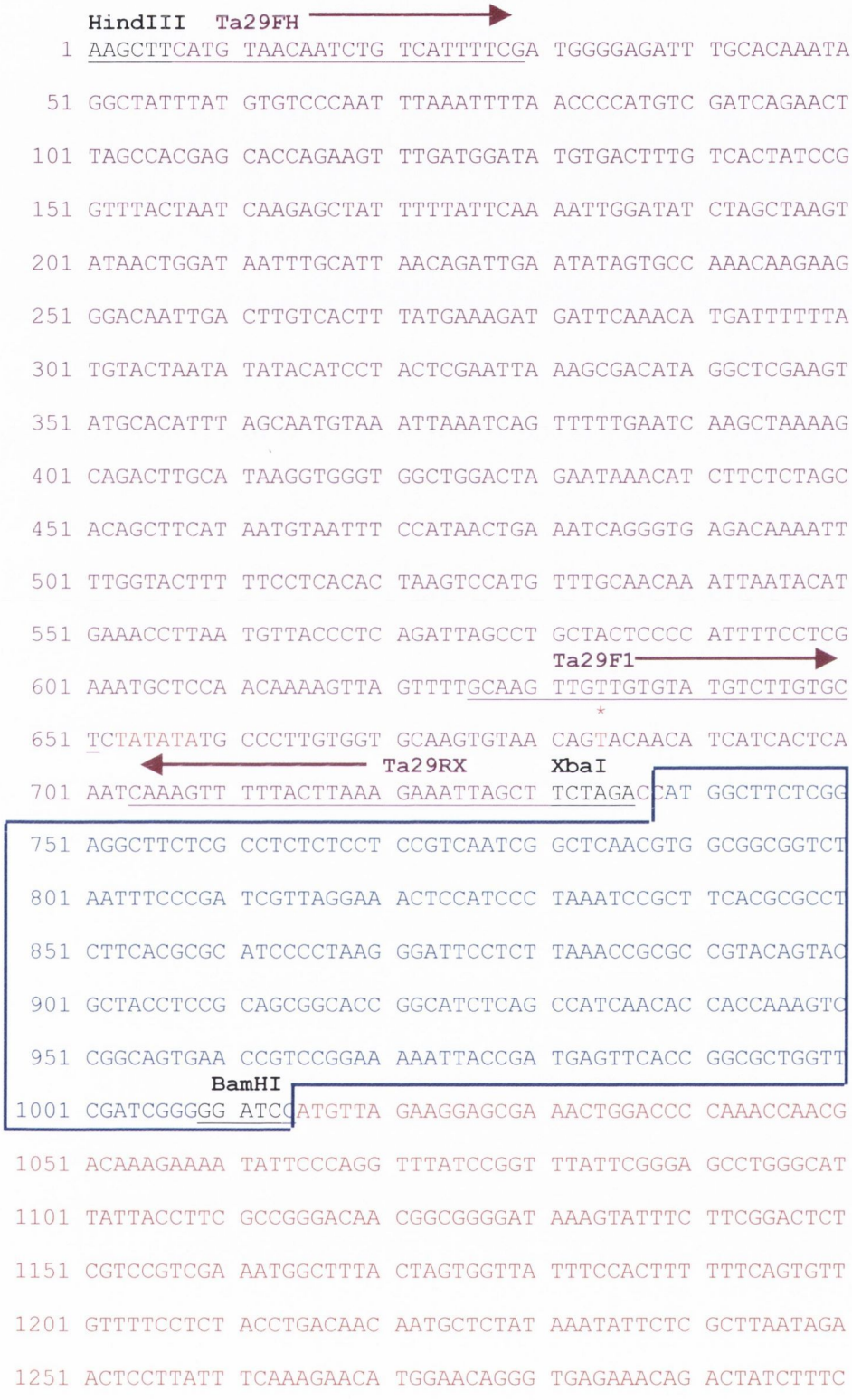


Figure 5.19 PCR product and restriction digests confirming construction of pBin-Ta29- βsubunitorfC9 constructs.

A. PCR of the Ta29 pollen specific promoter of 720 bp size in *N. tabacum,* lane 1 is the 1kb marker. **B.** Restriction digestion of the pRok219-35S constructs; pRok219-35S-*β*90-orfC9 (lane 5), pRok219-35S-*β*60-orfC9 (lane 8) and pRok219-35S-orfC9 (lane 11) with *Xba*I/*Hind*III. Undigested pRok219-Ta29-*β*90-orfC9 (lane 4), pRok219-Ta29-*β*60-orfC9 (lane 7), and pRok219-Ta29-orfC9 (lane 10). The 1 kb marker is in lanes 3, 6 and 9. **C.** Restriction digestion of pRoK219-Ta29 constructs; pRok219-Ta29-*β*90-orfC9 (lanes 2 and 3), pRok219-Ta29-*β*90-orfC9 (lane 5 and 6), restriction digestion of pRok219-orfC9 with with *Xba*I/*Hind*III (lanes 8 and 9). Lanes 1, 4 and lane 7 is the 1 kb ladder. **D.** Restriction digestion with *Hind*III/*Eco*RI of the three binary vector constructs isolated from *Agrobacterium* LBA4404; 1.7 kb pBin19-Ta29-*β*90-orfC9 (lane 4), 1.6 kb pBin19-Ta29-*β*60-orfC9 (lane 3), and 1.4 kb pBin19-Ta29-orfC9 (lane 2). The 1 kb ladder is in lane 1.

A.


HindIII
Ta29FH
1 AAGCTTCATG TAACAATCTG TCATTTTCGA TGGGGAGATT TGCACAAATA
51 GGCTATTTAT GTGTCCCAAT TTAAATTTTA ACCCCATGTC GATCAGAACT
101 TAGCCACGAG CACCAGAAGT TTGATGGATA TGTGACTTTG TCACTATCCG
151 GTTTACTAAT CAAGAGCTAT TTTTATTCAA AATTGGATAT CTAGCTAAGT
201 ATAACTGGAT AATTTGCATT AACAGATTGA ATATAGTGCC AAACAAGAAG
251 GGACAATTGA CTTGTCACTT TATGAAAGAT GATTCAAACA TGATTTTTTA
301 TGTACTAATA TATACATCCT ACTCGAATTA AAGCGACATA GGCTCGAAGT
351 ATGCACATTT AGCAATGTAA ATTAAATCAG TTTTTGAATC AAGCTAAAAG
401 CAGACTTGCA TAAGGTGGGT GGCTGGACTA GAATAAACAT CTTCTCTAGC
451 ACAGCTTCAT AATGTAATTT CCATAACTGA AATCAGGGTG AGACAAAATT
501 TTGGTACTTT TTCCTCACAC TAAGTCCATG TTTGCAACAA ATTAATACAT
551 GAAACCTTAA TGTTACCCTC AGATTAGCCT GCTACTCCCC ATTTTCCTCG
Ta29F1
601 AAATGCTCCA ACAAAAGTTA GTTTTGCAAG TTGTTGTGTA TGTCTTGTGC
651 TCTATATATG CCCTTGTGGT GCAAGTGTAA CAGTACAACA TCATCACTCA
Ta29RX
XbaI
701 AATCAAAGTT TTTACTTAAA GAAATTAGCT TCTAGACCAT GGCTTCTCGG
751 AGGCTTCTCG CCTCTCTCCT CCGTCAATCG GCTCAACGTG GCGGCGGTCT
801 AATTTCCCGA TCGTTAGGAA ACTCCATCCC TAAATCCGCT TCACGCGCCT
851 CTTCACGCGC ATCCCCTAAG GGATTCCTCT TAAACCGCGC CGTACAGTAC
901 GCTACCTCCG CAGCGGCACC GGCATCTCAG CCATCAACAC CACCAAAGTC
951 CGGCAGTGAA CCGTCCGGAA AAATTACCGA TGAGTTCACC GGCGCTGGTT
BamHI
1001 CGATCGGGGG ATCCATGTTA GAAGGAGCGA AACTGGACCC CAAACCAACG
1051 ACAAAGAAAA TATTCCCAGG TTTATCCGGT TTATTCGGGA GCCTGGGCAT
1101 TATTACCTTC GCCGGGACAA CGGCGGGGAT AAAGTATTTC TTCGGACTCT
1151 CGTCCGTCGA AATGGCTTTA CTAGTGGTTA TTTCCACTTT TTTCAGTGTT
1201 GTTTTCCTCT ACCTGACAAC AATGCTCTAT AAATATTCTC GCTTAATAGA
1251 ACTCCTTATT TCAAAGAACA TGGAACAGGG TGAGAAACAG ACTATCTTTC

```
1301 ACGAGTTGGA ATTTTCCCTA TATCGTGGGG TCAAACTCAC AGTAAAGGGA
                                                  KpnI
1351 CGAGACCAAC TCAAGGGAGA AACAACGCAT GAATAAGGTA CCGAGCTCGA
1401 ATTTCCCCGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT
1451 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
1501 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG
1551 TTTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC
1601 AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG TGTCATCTAT
                                  EcoRI
1651 GTTACTATGT TACTAGATCG GGAATTC
```

B.

```
    HindIII  Ta29HF
   1 AAGCTTCATG TAACAATCTG TCATTTTCGA TGGGGAGATT TGCACAAATA
  51 GGCTATTTAT GTGTCCCAAT TTAAATTTTA ACCCCATGTC GATCAGAACT
 101 TAGCCACGAG CACCAGAAGT TTGATGGATA TGTGACTTTG TCACTATCCG
 151 GTTTACTAAT CAAGAGCTAT TTTTATTCAA AATTGGATAT CTAGCTAAGT
 201 ATAACTGGAT AATTTGCATT AACAGATTGA ATATAGTGCC AAACAAGAAG
 251 GGACAATTGA CTTGTCACTT TATGAAAGAT GATTCAAACA TGATTTTTTA
 301 TGTACTAATA TATACATCCT ACTCGAATTA AAGCGACATA GGCTCGAAGT
 351 ATGCACATTT AGCAATGTAA ATTAAATCAG TTTTTGAATC AAGCTAAAAG
 401 CAGACTTGCA TAAGGTGGGT GGCTGGACTA GAATAAACAT CTTCTCTAGC
 451 ACAGCTTCAT AATGTAATTT CCATAACTGA AATCAGGGTG AGACAAAATT
 501 TTGGTACTTT TTCCTCACAC TAAGTCCATG TTTGCAACAA ATTAATACAT
 551 GAAACCTTAA TGTTACCCTC AGATTAGCCT GCTACTCCCC ATTTTCCTCG
                                        Ta29F1
 601 AAATGCTCCA ACAAAAGTTA GTTTTGCAAG TTGTTGTGTA TGTCTTGTGC
                                        *
 651 TCTATATATG CCCTTGTGGT GCAAGTGTAA CAGTACAACA TCATCACTCA
                                  XbaI
 701 AATCAAAGTT TTTACTTAAA GAAATTAGCT TCTAGACCAT GGCTTCTCGG
 751 AGGCTTCTCG CCTCTCTCCT CCGTCAATCG GCTCAACGTG GCGGCGGTCT
 801 AATTTCCCGA TCGTTAGGAA ACTCCATCCC TAAATCCGCT TCACGCGCCT
 851 CTTCACGCGC ATCCCCTAAG GGATTCCTCT TAAACCGCGC CGTACAGTAC
 901 GCTACCTCCG CAGCGGCAGG ATCCATGTTA GAAGGAGCGA AACTGGACCC
```

951 CAAACCAACG ACAAAGAAAA TATTCCCAGG TTTATCCGGT TTATTCGGGA

1001 GCCTGGGCAT TATTACCTTC GCCGGGACAA CGGCGGGGAT AAAGTATTTC

1051 TTCGGACTCT CGTCCGTCGA AATGGCTTTA CTAGTGGTTA TTTCCACTTT

1101 TTTCAGTGTT GTTTTCCTCT ACCTGACAAC AATGCTCTAT AAATATTCTC

1151 GCTTAATAGA ACTCCTTATT TCAAAGAACA TGGAACAGGG TGAGAAACAG

1201 ACTATCTTTC ACGAGTTGGA ATTTTCCCTA TATCGTGGGG TCAAACTCAC

KpnI

1251 AGTAAAGGGA CGAGACCAAC TCAAGGGAGA AACAACGCAT GAATAAGGTA

1301 CCGAGCTCGA ATTTCCCCGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT

1351 AAGATTGAAT CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT

1401 TGAATTACGT TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT

1451 ATGAGATGGG TTTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC

1501 GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG

EcoRI

1551 TGTCATCTAT GTTACTATGT TACTAGATCG GGAATTC

C.

HindIII Ta29HF

1 AAGCTTCATG TAACAATCTG TCATTTTCGA TGGGGAGATT TGCACAAATA

51 GGCTATTTAT GTGTCCCAAT TTAAATTTTA ACCCCATGTC GATCAGAACT

101 TAGCCACGAG CACCAGAAGT TTGATGGATA TGTGACTTTG TCACTATCCG

151 GTTTACTAAT CAAGAGCTAT TTTTATTCAA AATTGGATAT CTAGCTAAGT

201 ATAACTGGAT AATTTGCATT AACAGATTGA ATATAGTGCC AAACAAGAAG

251 GGACAATTGA CTTGTCACTT TATGAAAGAT GATTCAAACA TGATTTTTTA

301 TGTACTAATA TATACATCCT ACTCGAATTA AAGCGACATA GGCTCGAAGT

351 ATGCACATTT AGCAATGTAA ATTAAATCAG TTTTTGAATC AAGCTAAAAG

401 CAGACTTGCA TAAGGTGGGT GGCTGGACTA GAATAAACAT CTTCTCTAGC

451 ACAGCTTCAT AATGTAATTT CCATAACTGA AATCAGGGTG AGACAAAATT

501 TTGGTACTTT TTCCTCACAC TAAGTCCATG TTTGCAACAA ATTAATACAT

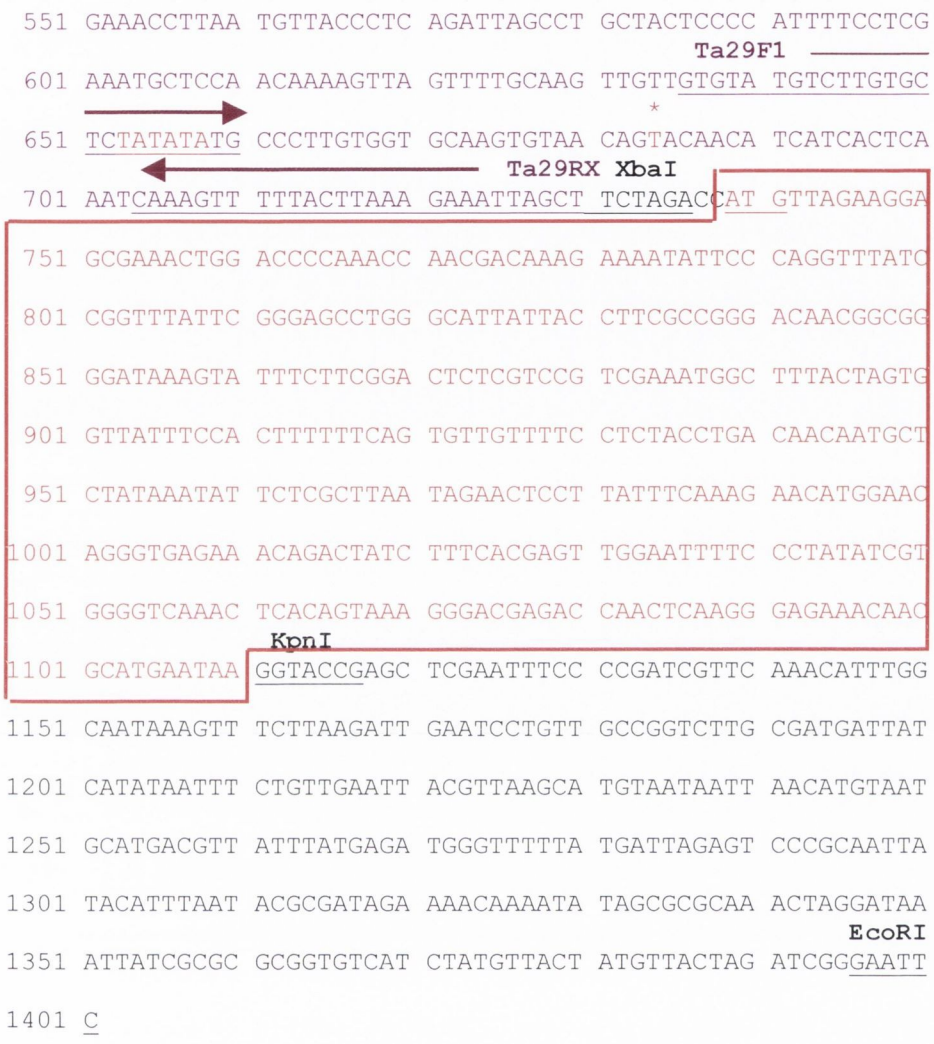


Figure 5.20 Confirmation of the sequence within the *Hind*III/*Eco*RI restriction sites for the three constructs in the binary vector pBin19.

A. is the pBin19-Ta29-*ß*90-orfC9 constuct, **B.** the pBin19-Ta29-*ß*60-orfC9 construct, and **C.** is the pBin19-Ta29-*ß*90-orfC9 construct. The purple colour type indicates the sequence for Ta29 promoter, the blue color type is for the trans-peptide (*ß*-90 or *ß*-60) sequence, the red color type is for the orfC9 sequence, and the black color type is for the NosT terminator sequence. The restriction sites for *Bam*HI, *Kpn*I, *Xba*I, are highlighted. Primers Ta29FH, Ta29RX are used in the amplification of 736bp size Ta29 fragment. The transcriptional start site is indicated by an aesterick * at position 684 bp, the putative TATA signal box is at position 653 bp (Seurinck *et al.*, 1990).

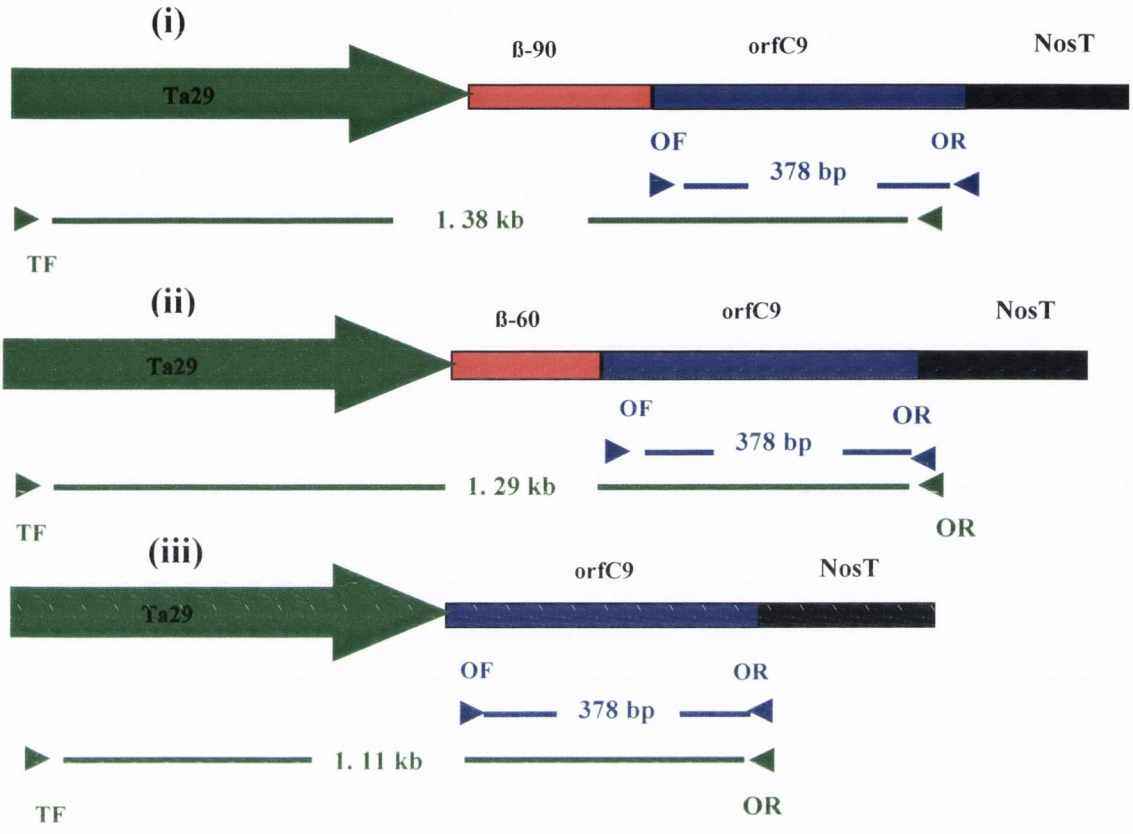


Figure 5.21 PCR analysis confirming the insert for the three different transgenes. The three different transgene constructs (**i, ii, iii**) under transcriptional control of the Ta29 promoter, the pre-sequence *ß*-60 or *ß*-90 , the CMS associated orfC9 gene and the NosT terminator are shown. PCR product using the **TF** and **OR** primer set amplified fragment sizes of 1.38 kb, 1.29 kb, 1.11 kb in the three constructs represented by (**i, ii, iii**). The orfC9 fragment of 378 bp is amplified in the three different transgenes, and the primers used for their amplification is the **OF** and **OR** primer set.

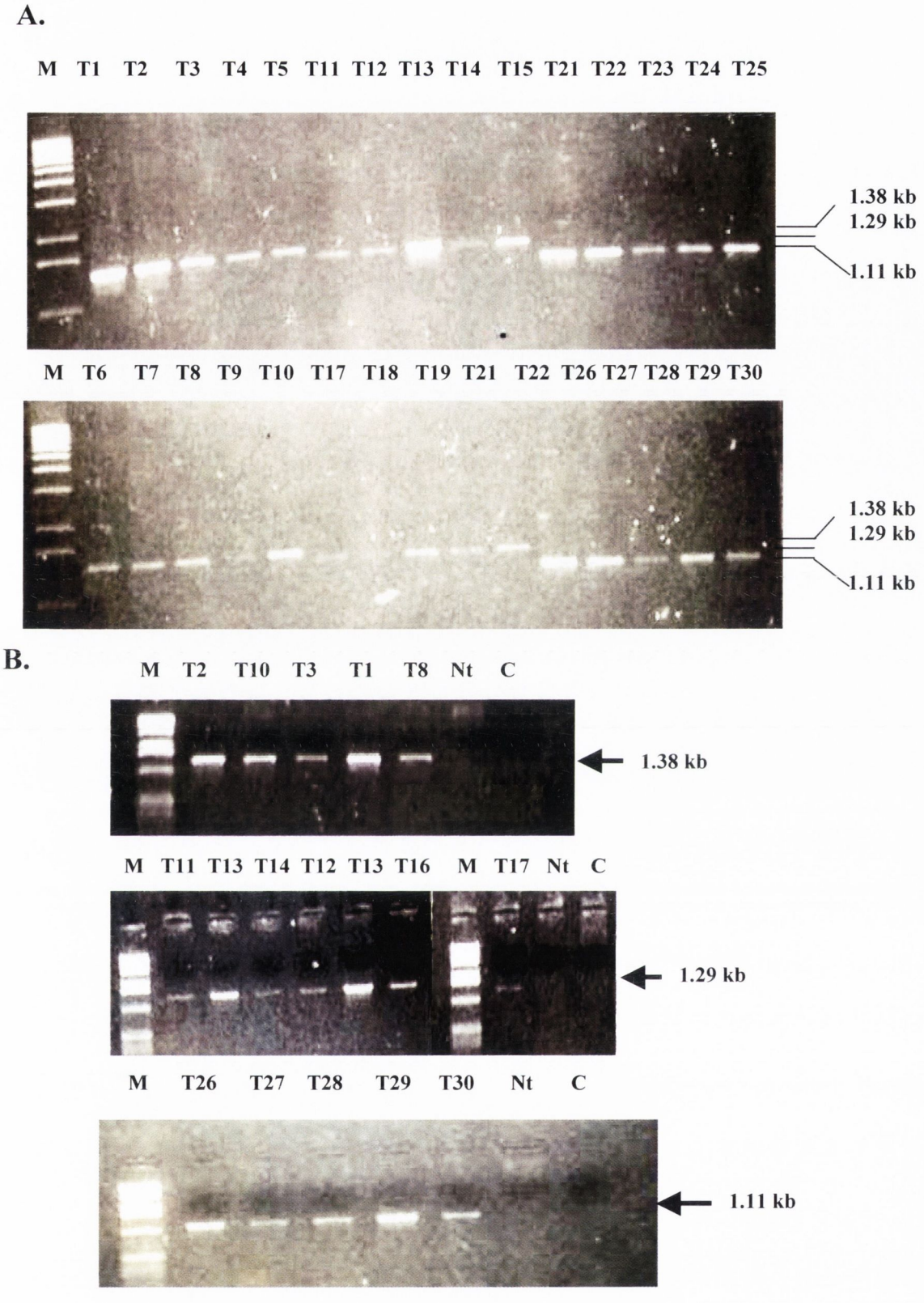


Figure 5.22 PCR confirmation of the Ta29-ßsubunit-orfC9 transformant.

A. Indicates the three different constructs for the transformed lines analyised by PCR using the primer TF and OR. The lines T1 to T10 refers to the Ta29-β90-orfC9 transformant, lines T11 to T20 refers to the Ta29-β60-orfC9 transformant, lines T21 to T30 refers to the Ta29-orfC9 transformant, and all these lines were positive by PCR. **B.** Illustrates PCR analysis for some of the transformed lines (Tn), that includes the negative controls; non-transformed tobacco plants (Nt), and minus DNA template (−). Lane M refers to the 1 kb marker.

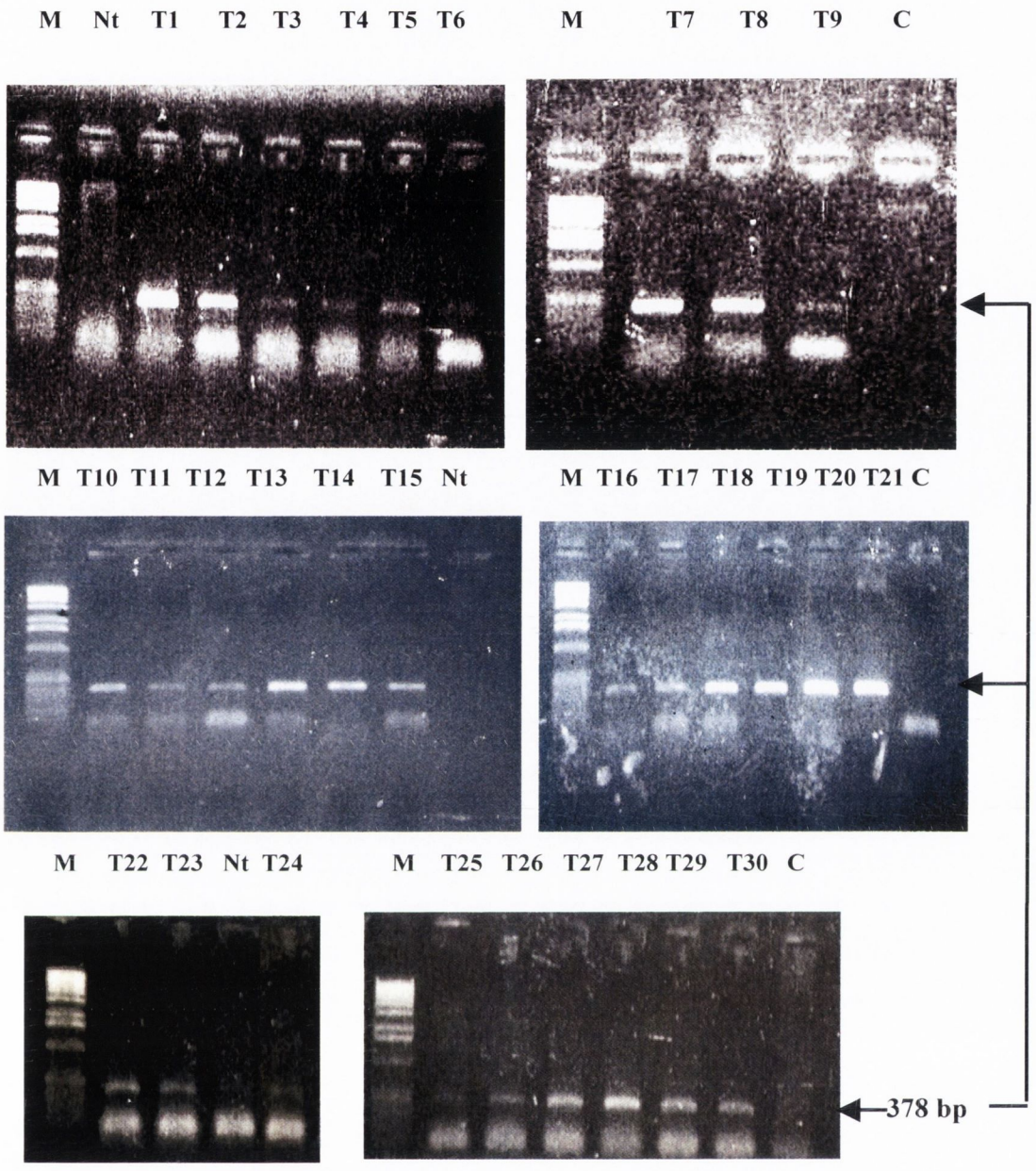


Figure 5.23 PCR analysis of Ta29-β90-orfC9, Ta29-β60-orfC9, Ta29-orfC9 using the primer set orfC9F/orfC9R.

The Total DNA extracted from the T1 to T30 transformed plants lines were analysised. by PCR. A PCR product of 378 bp in size was amplified, and only the transformed plants that gave a positive PCR product are shown. Two negative controls were used here, a wild type control (Nt) and a minus DNA template control (C). Lane M refers to the 1 kb marker.

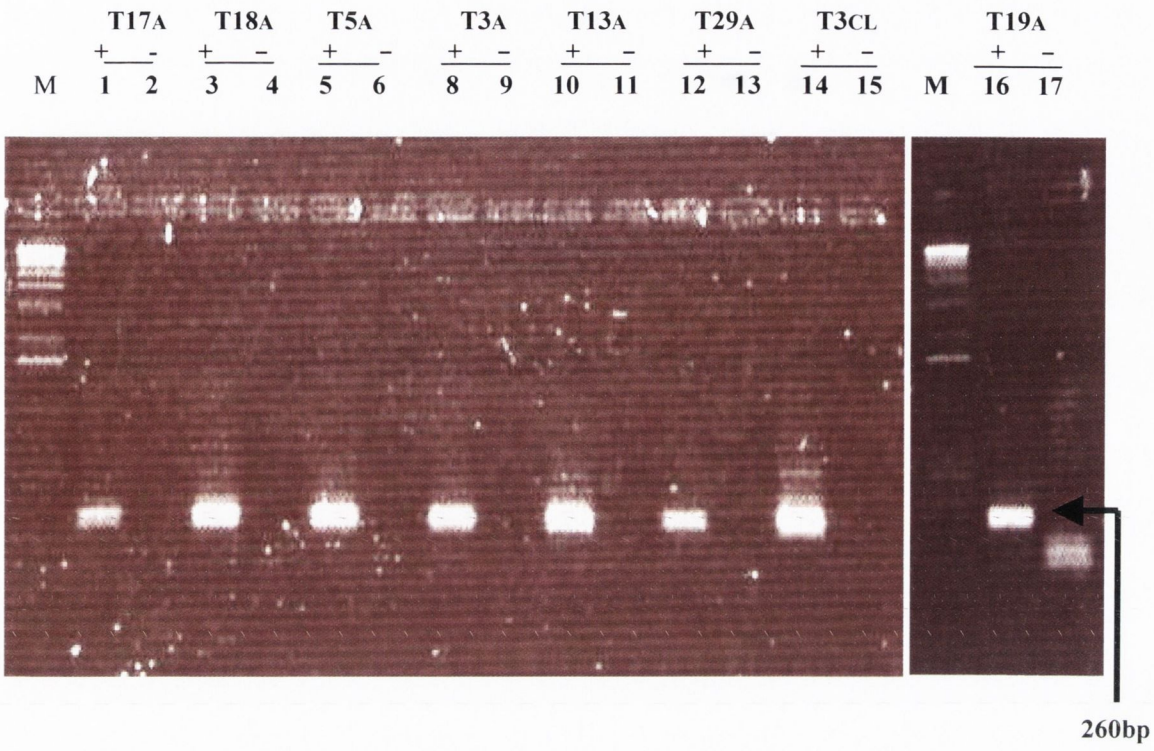


Figure 5.24 The Presence of transcripts by RT-PCR in anthers from transformed lines containing the Ta29 promoter.

PCR reaction showing the amplification of a 260 bp sized product with the orfC9 primer set (orfc9intF and orfC9intR). cDNA from anther tissue (A) for the **To** transformed lines of T3, T5, T13, T17, T18, T19, T29 and cDNA from leaf tissue (L) of the T3 CaMV35S transformed line – T3CL is used as template. Lane M refers to the 1 kb marker.

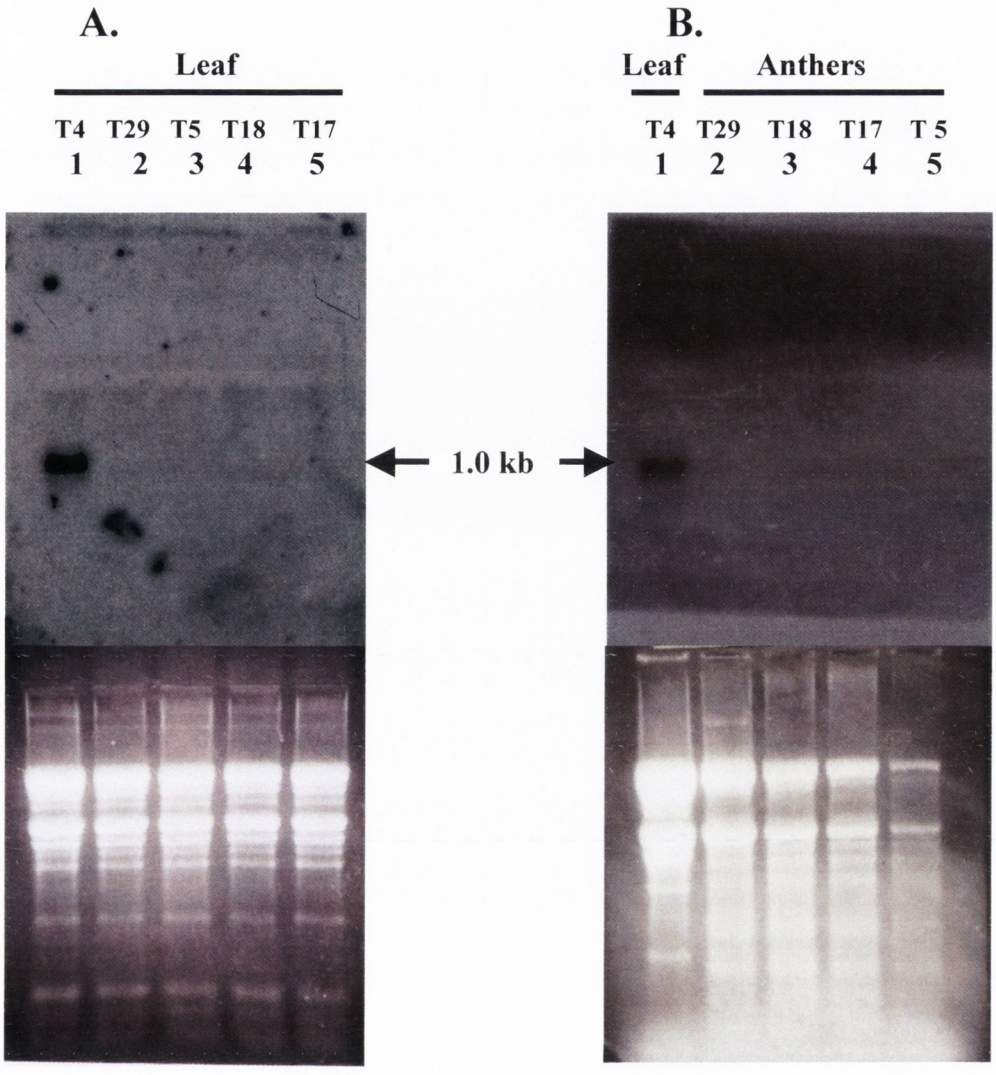


Figure 5.25 Northern Blot analysis of Total RNA from anther tissue in the Ta29 constructs.

A. Is a control experiment with total RNA from leaf tissue from the Ta29 transformants (Ta29-*ß*90-orfC9, Ta29-*ß*60o-rfC9, Ta29-orfC9) in the lines T5, T17, T18, T29 and in the CaMV35S transformant line T4 (positive control) were probed with orfC9. No transcripts were detected in the Ta29 transformant lines but a 1.0 kb size transcripts was present in the leaf tissue for CaMV35S transformant line T4. **B.** Total RNA from anther tissue in the Ta29 transformant lines T5, T17, T18, T29 was probed with orfC9, no transcripts were detected except for a 1.0 kb size transcript in the total RNA from leaf tissue for the CaMV35S transformant line T4.

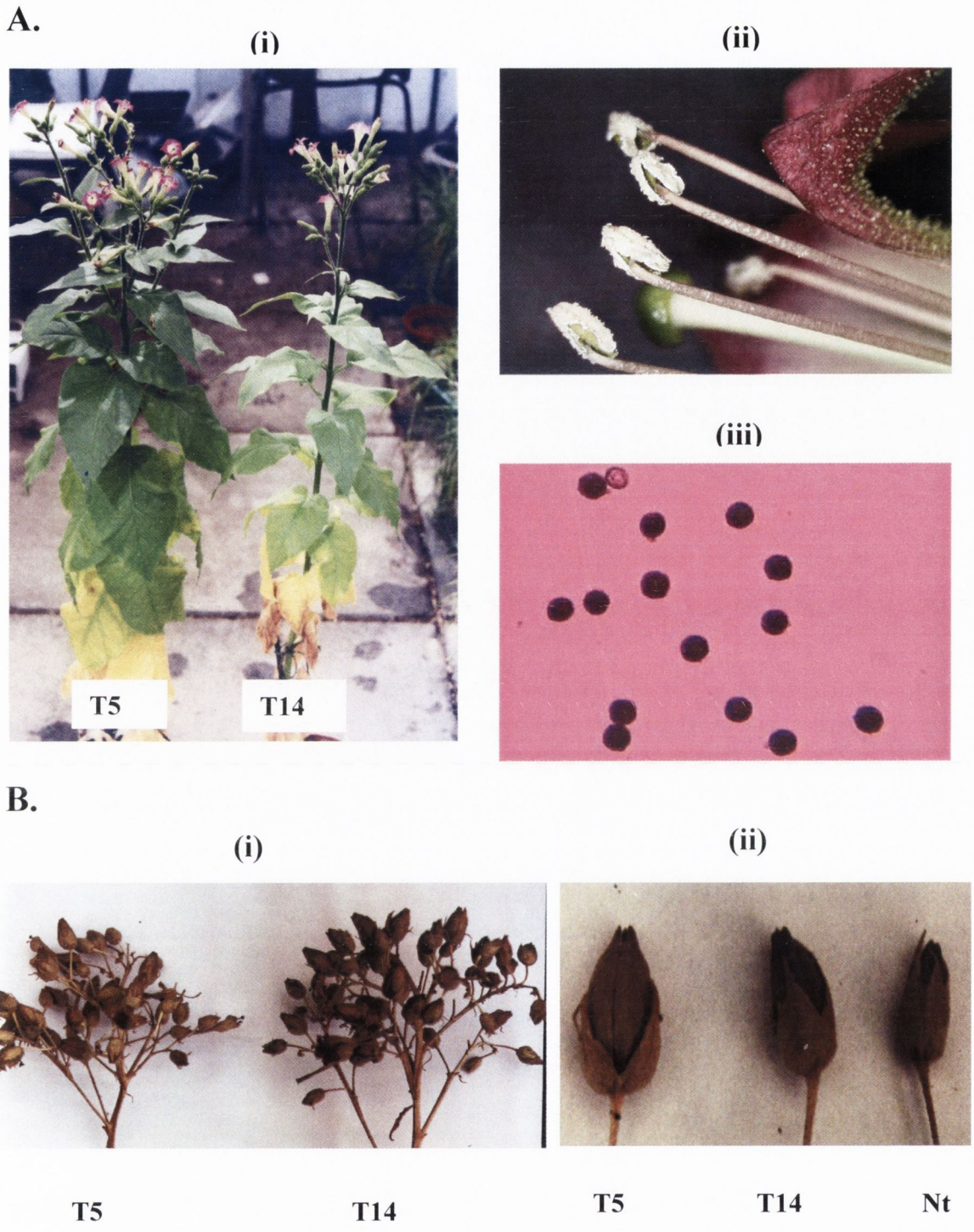


Figure 5.26 Flower Morphology of Ta29 Transformant plant lines in their To generation. A. (i) Two of the transformed lines (T5, T14) grown to flowering. **(ii)** A transformed plant line Ta29-*ß*90-orfC9 showing a fertile phenotype, anthers located well above the stigma and the golden brown pollen grains. **(iii)** The pollen grains are fully stained with Lugol solution (Potassium Iodide). **B. (i)** Fruit capsule for two seperate transformant lines (T5, T14) with a normal fertile seed set. **(ii)** Closer examination of fruit capsule from transformed plants lines (T5, T14) and non transformed line (Nt).

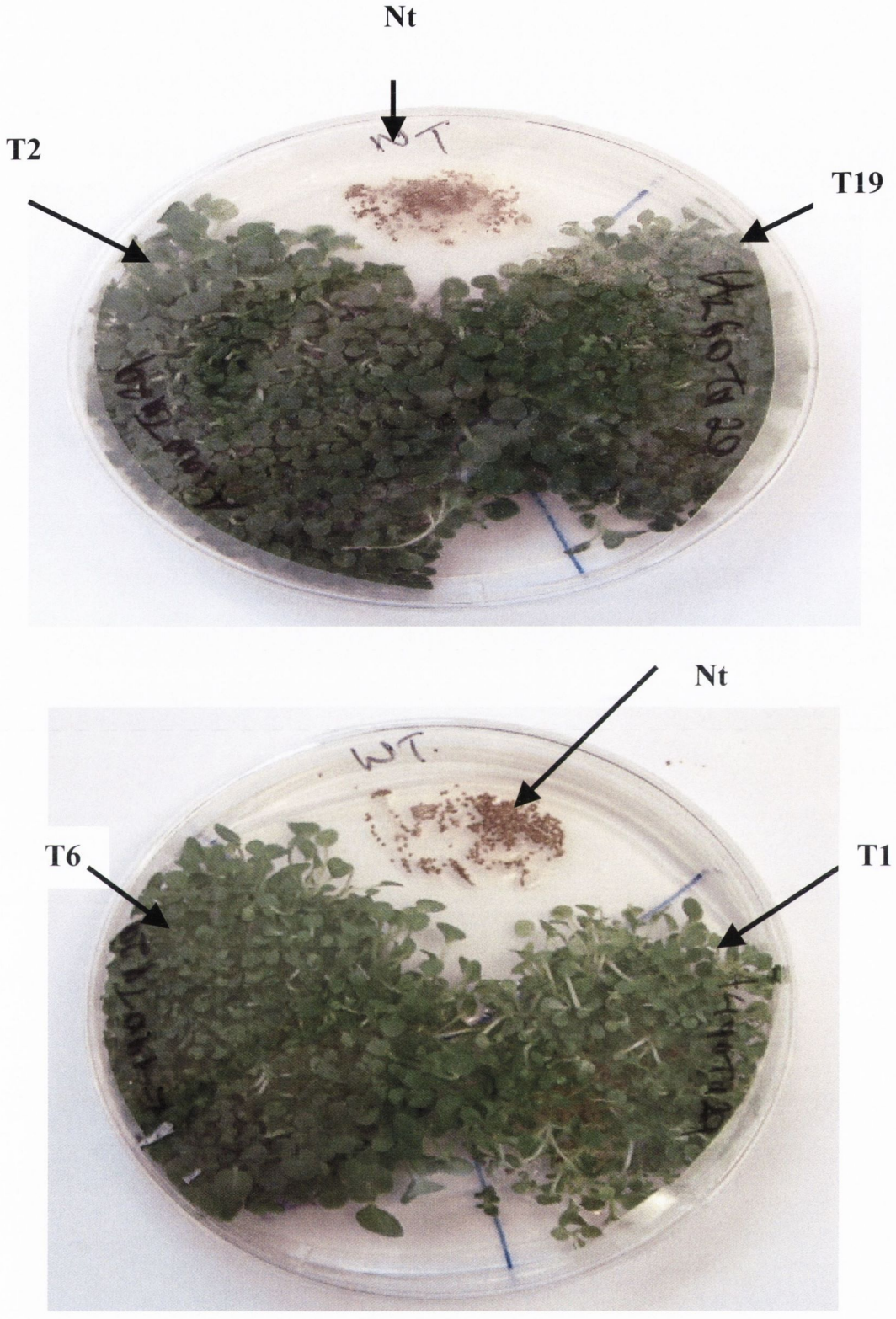


Figure 5.27 Illustrates Ta29 promoter seedlings capable of germinating on MS Kanamycin (500 µg/ml)medium.

The transgenic lines T1, T2, T6 and T19 from the **To** generation is an example of Ta29 transgene lines capable of germinating on MS Kanamycin (500 µg/ml) medium.

Table 5.1 Phenotypic characteristic of CaMV35S transgenic plants To generation

Plant Lines	No. of plants	Classification of Fertility			Number of pollen per Anther (10^4)	% Pollen fertility (F) per Flower		Seeds per capsule (mg)
		F	S/S	S		F	S	
Nt	4	4	–	–	22.0 ± 3.5	95 ± 1.75	5 ± 1.75	140 ± 12.2
35S-β90-orfC9	10	8	–	–	17.2 ± 4.9	92.9 ± 2.5	7.1 ± 2.5	126 ± 23.0
		–	T3	–	4.6 ± 1.13	40.4 ± 4.9	59.5 ± 4.9	32.5 ± 10.6
			T4	–	13.8 ± 1.7	60.0 ± 2.0	40.0 ± 2.0	12.5 ± 8.2
35S-β60-orfC9	10	10	–	–	14.4±1.9	89.7± 2.2	10.3 ± 2.2	153.3 ± 26.5
35S-orfC9	10	10	–	–	15.8±3.0	92.2±3.4	7.8 ± 3.4	154.0 ± 25.0
					$\bar{x} \pm sd$			

Non-transformed-Nt, Fertile-F, Sterile-S, Semi-sterile-S/S

$\bar{x}$ = average value (pollen grain number, pollen fertility/sterility (%) or seed set mass (mg))

sd = standard deviation value

Table 5.2 Segregation of T_1 seeds from a series of Transgenic lines onto MS Kanamycin (500 µg/ml) medium.

Approximately 100 seeds were plated (except for line T3 andT4 where only approximately 50 seeds were used) in duplicates onto the MS medium; 5 different lines for the 35S-β90-orfC9 transformant (T2, T3, T4, T5, T7), 2 different lines for the 35S-β60-orfC9 transformant (T14 and T15), and 4 different lines for the 35S-orfC9 transformant (T20, T21, T25, T27).

Line Tested	Seed Number	Germinated	Kan + (Green)	Kan – (Bleached)	Ratio
T2	100	80	67	13	**5 : 1**
T3	50	36	46	0	–
T4	40	24	24	0	–
T5	100	71	56	15	**3.7 : 1**
T7	100	59	59	0	–
T14	100	78	78	0	–
T15	100	62	62	0	–
T20	100	60	60	0	–
T21	100	54	42	12	**3.7 : 1**
T25	100	49	41	8	**5 : 1**
T27	100	53	48	5	**9.6 : 1**

Table 5.3 Phenotypic characteristic of CaMV35S transgenic plants T1 generation

Plant Lines	No. of plants	Plant Fertility F	S/S	S	Pollen/Anther No. (10^4)	Pollen/Flower, % F	S	Seed/capsule (mg)
Nt	3	3	–	–	24.1 ± 2.4	96.0± 3.1	4.0± 3.1	136.0 ± 9.2
35S-β90-orfC9	6	–	T3 (3)	–	8. 8 ± 1.2	48.0 ± 1.2	52.0 ± 1.2	13.3 ± 9.2
			T4 (3)	–	9. 5 ± 3.1	65.0 ± 2.2	35.0 ± 2.2	19.2 ± 6.3
35S-orfC9	3	3	–	–	14.2 ± 1.9	91.7± 5.2	9.3 ± 5.2	122.0 ± 10.2
					$\bar{x} \pm sd$			

Non-transformed-Nt, Fertile-F, Sterile-S, Semi-sterile-S/S

$\bar{x}$ = average value (pollen grain number, pollen fertility/sterility (%) or seed set mass (mg))

sd = standard deviation value

Table 5.4: Phenotypic characteristics of Ta29 transgenic plants in To generation

Plant Lines	No. of plant	Classification of Fertility			Number of pollen cells per Anther (10^5)	% Pollen fertility (F) per Flower		Seeds per capsule (mg)
		F	S/S	S		F	S	
Nt	4	4	–	–	2.5 ± 3.5	95.0 ± 2.1	5.0 ± 2.1	135.0 ± 13. 2
Ta-β90-orfC9	10	10	–	–	2.2 ± 0.38	91.8 ± 1.7	8.2 ± 1.7	122.0 ± 42. 5
Ta-β60-orfC9	10	10	–	–	3.24 ± 1.92	93.7 ± 2 1	6.3 ± 2.1	111.6 ± 12. 5
Ta-orfC9	10	10	–	–	1.97 ± 0.35	96.0 ± 1 3	4.0 ± 1 3	138.8 ± 36.3
					$\bar{x} \pm sd$			

Non-transformed-Nt, Fertile-F, Sterile-S, Semi-sterile-S/S

$\bar{x}$ = average value (pollen grain number, pollen fertility/sterility (%) or seed set mass (mg))

sd = standard deviation value

CHAPTER 6

Pollen Development in Fertile and CMS *L. perenne*

6.1 Introduction

6.1.1 An overview of the process of pollen development in higher plants

The male reproductive organ in flowering plants is the stamen which consists of an anther attached to a stalk-like filament. The production of the male gametophytes, pollen grains, occurs in the anther locules and is initiated from a group of diploid sporogenous pollen mother cells (PMCs) (Mascarenhas, 1990; Bedinger, 1992; McCormick, 1993). Each anther is typically sub-divided into four compartments or locules, the walls of which comprise 3 discrete cell layers: epidermis, middle layer and endothecium. A fourth layer, the tapetum, surrounds the PMCs and is intimately involved in the developmental progression from PMC to mature pollen grain.

Pollen development has been shown to proceed in two phases (Koultunow *et al.*, 1990; Bedinger, 1992; Goldberg *et al.*, 1993). At the beginning of Phase I sporogenous cells are generated. These cells differentiate to form the pollen mother cells (PMCs) or meiocytes. PMCs are separated from other anther cells by being encased in a callose (β-1,3-glucan) wall. PMCs then undergo two meiotic divisions to produce callose-encased tetrads which contain four haploid microspores. During this stage the microspores are separated from the nutritive fluid in the anther locule by callose. Young microspores are ultimately released into the anther locule by β-1,3-glucanase-mediated digestion of callose, which marks the start of the second phase of pollen development, during which the outer pollen wall (exine) is constructed.

During Phase II, the single haploid nucleus migrates towards the cell periphery (opposite to the pollen pore), a process termed polarization. The microspore cytoplasm begins to fill with small vacuoles which then converge to form a single large vacuole that compresses the cytoplasm into a small region

directly opposite the cell pore. The microspore then undergoes mitosis to produce a generative (sperm) nucleus and a vegetative nucleus which become surrounded by cytoplasm to form individual cells. The vegetative cell is small and rich in plastids, mitochondria, lipid bodies, starch grains and small vacuoles. The larger generative cell contains a nucleus, mitochondria and vacuoles (Fei and Swahney, 2001). This binuclear or bicellular product of microspore mitosis is essentially the young pollen grain. In *Z. mays* (Warmke and Lee, 1977), the generative cell undergoes a second mitotic division to form two sperm cells. However in *Arabidopsis,* the second mitotic division occurs after germination of the pollen tube (Palmer *at al.*, 1992; Fei and Swahney, 2001).

The pollen grain enters a period of maturation marked by development of the inner pectin wall (intine), the filling of the large central vacuole with starch granules to be used as a source of energy during germination, and finally dehydration during which its water content is reduced by up to 60%.

6.1.2 The role of the tapetum during pollen development

The tapetum is the cell layer that lines the anther locule where pollen development and maturation takes place (Bedinger, 1992). The production of mature pollen grains requires involvement of the haploid gametophytic cells and sporophytic diploid cells. The formation and release of pollen requires the co-ordinated expression of a complex array of genes in both the developing gametophyte and the sporophytic tissues of the anthers, most notably the tapetum (Pacini *et al.*, 1985). During the early stages of meiosis, tapetal cells become binucleate and dramatically differentiate into polar secretory cells (Warmke and Lee 1977; Wu and Cheung, 2000). The mature tapetal cells lack primary cell walls, and the secretory cells are tightly packed on the tapetal surface, facing out into the locule. Electron micrographs of tapetal cells reveal the presence of mitochondria, ribosomes, endoplasmic reticulum, Golgi apparatus, and many secretory vesicles located on the locular surface (Bedinger, 1992; Papini *et al.*, 1999; Suzuki *et al.*, 2001). The tapetal layer in the majority of plants remains associated with the anther wall and perform a sceretory role during pollen development e.g in maize, wheat, rice and *Arabidopsis* (Warmke and Lee, 1977; Wang *et al.*, 2003; Ku *et al.*, 2003; Regan and Moffatt, 1990). Earlier studies

indicated a nutrititive role for the tapetum during microspore development, very similar to the role of nurse cells in mammalian species (Hesse and Hess, 1993).

The tapetum also secretes the β-1,3-glucanase that digests callose to release individual haploid microspores from tetrads (Steiglitz, 1977; Rhee and Somerville, 1998). The timing and temporal regulation of β-1,3-glucanase secretion is critical for normal pollen development. CMS in *Petunia* has been attributed to premature digestion of the callose wall (Izhar and Frankel, 1971) and in tobacco, plants have been engineered to produce a male-sterile phenotype by fusing a callase transgene (β-1,3-glucanase) to promoters that are active in the tapetum (Worrall *et al.*, 1992; Tsuchiya *et al.*, 1995). Meiotic events occurred normally in the absence of callose but the pollen wall developed abnormally due to aberrant exine deposition.

Shortly after microspores become vacuolated and are released from the tetrad, the tapetum begin to degenerate. This is signified by the polarization of cytoplasmic materials, vacuolation, thinning of the tapetal cell walls and cell shrinkage. Finally, the cells rupture and release their cellular contents some of which, e.g. tryphine and pollenkit, contribute to pollen wall development. Other substances are proteinaceous and lipid-like in nature and are required for the interaction between pollen and stigma, and also in the formation of a scaffold upon which other molecules may be deposited (Chapmann, 1987; Bedinger, 1992; Wu *et al.*, 1997; Piffanelli and Murphy, 1998). The developmentally scheduled death and disintegration of the tapetal cell layer shares some similarities with programmed cell death in animals, for example, oligonucleosomal cleavage of DNA and release of *cytochrome c* (Balk and Leaver, 2001) (as discussed in Chapter 1). However, it is still unclear whether the two processes are mechanistically similar at the molecular level.

6.1.3 Pollen development and mutations resulting in a male-sterile phenotype

A characteristic feature of pollen grains is the outer layer called the pollen wall. The wall is a complex and highly sculptured structure that plays an important role in protecting the pollen grain from a variety of environmental stresses and potential pathogens during pollination. The wall consists of two different layers: an inner layer called the intine which is located adjacent to the plasma membrane and an outer layer called the exine. The intine, which itself

comprises 2 layers (exintine and endintine) is composed of pectin, cellulose and proteinaceous materials secreted by the microspore during the vacuolation stage in the microspore (Owen and Makaroff, 1995). The exine is also composed of two layers, the innermost nexine (lying on top of the exintine) which is featureless and the outermost highly sculptured sexine that consists of multiple pores and furrows (Shukla *et al.*, 1998). The exine is primarily composed of sporopollenin, one of the most durable organic substances known, which serves as a rigid protective exoskeleton for the male gametophyte (Bedinger, 1992).

Arabidopsis is probably the most studied plant in terms of pollen wall development. There are several types of mutants in *Arabidopsis* that show defects in the formation of the exine which results in a male-sterile phenotype: *flp1* (*faceless pollen-1*), *nef1* (*no exine formation*), *ms1* (*male sterility-1*), *ms2* (*male sterility-2*), *dex1* (*defective in exine formation-1*) (Paxson-Sowders *et al.*, 2001; Ariizumi *et al.*, 2004; Barrena and Wilson 2006; Aarts *et al.*, 1997; Ariizumi *et al.*, 2003). The *flp1* mutant showed excessive tryphine covering the exine baculae (columns). Although the microspores and the exine appeared normal, with sporopollenin and bacula formation, on closer examination the exine was irregular with broken parts. The *flp1* mutation was therefore assumed to play an important role in sporopollenin biosynthesis, particularly during the later stages of pollen development. Abnormal walls were also found in the *nef1* mutant which exhibited defects in primexine formation and although sporopollenin was produced, it failed to properly polymerize on the outer layer of the microspore (Ariizumi *et al.*, 2004). In the *dex1* mutant, sporopollenin production was normal but primexine development and the deposition of sporopollenin was abnormal. It has been proposed that the DEX1 protein might be a component of primexine that plays a role in its polymerisation (Paxson-Sowders *et al.*, 2001). The exine layer was completely absent in the *ms2* mutant, suggesting a defect in sporopollenin deposition. The MS2 protein encodes a reductase which reduces long fatty acids to fatty alcohols that are required for sporopollenin production.

6.2 Chapter Aim

A comparative investigation of pollen development in fertile and CMS *L. perenne* plants was undertaken in order to identify the developmental stage at which pollen abortion occurred. This was accomplished by examination of thin sections of anthers by light microscopy.

6.3 Materials and Methods

6.3.1 Plant Material

L. perenne CMS9B290 (male-sterile) and LPS21 (fertile) were grown in well drained soil in a cool greenhouse. The lines were also kept in natural outdoor conditions in order to stimulate flowering.

6.3.2 Isolation of Anthers

The anthers were teased from the primary florets using a fine forceps and a scalpel blade. When viewing pollen grains in the anther at maturation, the anthers were either burst or cut open to release the pollen grains immediately into Lugol staining solution on a glass slide. The sample was then viewed by light microscopy.

6.3.3 Fixing and embedding anther tissue in epoxy araldite resin

Anthers were isolated and sorted according to their size and were immediately fixed in 4 ml primary fixative solution (3.0% gluteraldehyde (v/v) in 0.05 M potassium phosphate buffer (pH 6.8)) in a stoppered glass vial. The vials were placed in a rotator for gentle agitation and samples were allowed to fix for a period of 1.5 hr. at room temperature. The primary fixative solution was carefully removed and samples were washed 6 times (10 min per wash) in 0.05 M phosphate buffer with gentle agitation to remove unreacted glutaraldehyde. Samples were then incubated in 4 ml 2% osmium tetroxide (w/v) in 0.05 M phosphate buffer (pH 6.8) for 30 min in a fume hood with gentle agitation. The anthers were then washed in 10% (v/v) acetone for 10 minutes, and dehydrated by further washes in an increasing acetone series (30% acetone, 50% acetone, 70% acetone, 95% acetone) for 10 minutes each and finally in two separate 100% acetone washes for 30 minutes each. The 100% acetone solution was gradually replaced with epoxy resin. This was initially done by immersing the anther tissue in 100% propylene oxide solution for two separate wash stages of 15 min and 30 min, respectively. The final 100% propylene oxide was replaced with a 1:1 mixture of 50% propylene oxide and epoxy araldite (24.0 g 100 epoxy resin agar, 16.0 g DDSA hardener, 10.0 g MNA hardener, 1.5 g BDMA accelerator) and was allowed to mix by rotating for 3 hr. This infiltration was continued by transferring

the anthers to a fresh 100% epoxy araldite resin, and the glass tube with samples was agitated on the rotator for 3 hr. The anthers were transferred from the glass vials, using an applicator stick to a fresh 100% epoxy araldite in rubber beem capsules. The capsules containing the samples were placed in an oven under vacuum for 10 min. in order to remove any air bubbles from the epoxy araldite resin mix. The vacuum was gently released to ambient pressure, and the samples in the epoxy capsules were incubated at 60°C for 24 hr.

6.3.4 Light Microscopy

The resin embedded anthers were sliced into 3 micron sections using an LKB ultramicrotome and stained with 0.5% (w/v) toluidine blue containing 0.1% sodium carbonate, and were heat-fixed on a hotplate at 55-65°C. The anther tissue sections were then viewed using a Nikon light microscope.

6.4 Results

6.4.1 Pollen development in fertile and CMS *L. perenne* lines

The inflorescence in *L. perenne* is organized as a series of spikelets each of which contains approximately 9 closely appressed florets at slightly different stages of development, with the largest (oldest) floret located at the base and the smallest (youngest) located at the tip. By choosing infloresences that differ in age by ca. 7 days, and florets at different positions within a spikelet, it is possible to isolate a series of anthers of increasing length in which most stages of pollen development are represented (Figure 6.1). To examine the various stages anthers were fixed in glutaraldehyde and embedded in an epon araldite resin. Thin sections (3 micron) were prepared, stained with toluidine blue and examined by light microscopy.

I could distinguish 7 stages of anther/pollen development in *L. perenne* based on increasing anther size. These approximated the developmental stages identified by Warmke and Lee (1977) and Ku *et al.* (2003) in maize and rice, respectively. Each stage is described below:

Stage 1 - Pollen mother cells encased in callose: This is the earliest stage of pollen development that I observed. The four cell layers of the anther are well-defined: the epidermis, endothecium, middle layer, and the tapetum. The tapetum is clearly identifiable as a darkly-staining layer of uniformly rectangular cells surrounding a central mass of pollen mother cells (PMCs) or meiocytes. The PMCs, which contain a large prominent nucleus (n), are normally encased in callose at this stage. Here, they appear angular with spaces between them because during fixation or embedding the cells have become physically separated from each other. This conclusion is supported by the observation that the tapetum has also separated from the endothecium in both sections. In general, no morphological differences between sterile and fertile anthers were evident (Figure 6.2 A, B). This suggests that pollen development in the *L. perenne* CMS line is proceeding normally at this stage.

Stage 2 - Released PMCs: The PMCs stain more lightly and have become more rounded in appearance due to the dissolution of callose. In both fertile and sterile

anthers the PMCs appear similar in size with a prominent darkly stained nucleus visible in most cells (Figure 6.3 A, B). It is at this stage that meiosis occurs. Indeed, some of the nuclei in Figure 6.3 (B, C) have the appearance of nuclei at the prophase stage of meiosis. A comparable section at the same anther stage proved difficult to obtain for the fertile anthers. Small vacuoles are also visible in the PMCs. The tapetal layer is more fully developed and has increased in thickness in both fertile and sterile anthers (Figure 6.3B).

Stage 3 - Meiotic tetrads: The young microspores are organized as callose-encased tetrads indicating that they have completed meiosis. Small, now haploid, nuclei are visible in the PMCs in fertile and sterile anthers (Figure 6.4B). The tapetum (tp) in both the fertile and sterile sections appears more darkly stained (Figure 6.4 A, B). At this stage of anther development, the fertile and sterile anthers are indistinguishable.

Stage 4 - Young multi-vacuolated microspores: Following meiosis the callose wall surrounding the tetrads and dyads dissolves due to the action of β-1,3 glucanase resulting in release of the young microspores. In fertile anthers the cytoplasm of the youngest microspores is filled with tiny vacuoles (vp), and nuclei occupy a more or less central position (Figure 6.5). The microspores appear to lack walls because exine synthesis has not yet commenced. Also in fertile anthers, the tapetum remains very darkly stained with sharply defined cell walls and prominent nuclei. Unfortunately, it was not possible to obtain sections for this precise stage from sterile anthers.

Stage 5 - Vacuolated microspores: At this stage, in most of the microspores in both fertile and sterile anthers the vacuoles have coalesced and nuclei are being displaced to the periphery (Figure 6.6). The exine wall surrounding fertile and sterile microspores is also visible. In fertile anthers, the tapetal cell layer appears thicker but the cells have lost their rectangular appearance and smooth walls, and the nuclei are less apparent. The tapetal cells in the sterile anthers show evidence of disintegration and appear minutely vacuolated (Figure 6.6). The epidermal, middle and endothecium cell layers of the anther wall appear compressed in both

sterile and fertile anthers. Young pollen grains appear slightly smaller in sterile anther but otherwise similar to those in fertile anthers.

Stage 6 – Maturing pollen grains: Fertile anthers contain pollen grains with a single large off-centre vacuole and cell contents that accumulate near one pole (Figure 6.7). The tapetum is in a more advanced state of disintegration. In contrast, the pollen grains in sterile anthers have collapsed with the appearance of deflated footballs. The tapetal cells contain large vacuoles but retain a rectangular compressed appearance. The large darkly-stained mass may represent the released contents of tapetal cells (Figure 6.7).

Stage 7 – Mature anthers and pollen: The large vacuoles in some pollen grains in fertile anthers contain a densely staining material (starch), and in some grains a sickle-shaped vegetative cell is visible. The tapetum has almost completely disintegrated (Figure 6.8). In sterile anthers, pollen cell contents have almost completely disappeared (Figure 6.8).

Cross sections of entire mature anthers from fertile and sterile plants are shown in Figure 6.9. Anthers from fertile plants contain locules that possess several intact structural features. The section shows mature pollen grains with an outer exine (ex), a disintegrating tapetum, epidermal layer, endothecium, stomium and vascular bundles. The pollen grains are full and are located towards the periphery of the locule (Figure 6.9A). However, in the sterile line (Figure 6.9B), the tapetum is absent in all four anther locules and pollen grains appear mis-shapen and empty.

Mature pollen grains stained with Lugol solution are shown in Figure 6.10. Mature pollen grains removed from the surface of dehisced fertile anthers and placed in Lugol solution stain a deep uniform purple, revealing the high starch content of the cytoplasm. In contrast, sterile grains are smaller, misshapen, and contain only residual starch and cell contents.

6.5 Discussion

The overall purpose of this study was to examine the process of pollen development in *L. perenne* in order to identify the exact stage at which a noticeable difference occurred between fertile and CMS lines. Previous work by Connolly and Wright-Turner (1984) involved sectioning of mature anthers only with a view to classifying whether anthers were sterile or fertile (Connolly and Wright-Turner 1984). Here we show that the first 3 stages of pollen development i.e. up to and including meiosis, are indistinguishable between fertile and sterile plants and, moreover, that the first clear difference, vacuolation of tapetal cells, was observed at stage 5 in anthers of the CMS line. By stage 6, the development of microspores in sterile anthers was completely compromised. These findings mirror the development of pollen in fertile and sterile anthers in *Z. mays* (Warmke and Lee, 1977).

The cells of the tapetum in fertile anthers at Stage 3 and Stage 4 appear similar in size and in shape. The internal walls of the tapetum for both stages is smooth, centrally facing, with distinctive cell walls and with visible nuclei. However, by Stage 5, the edge of the tapetum facing towards the central anther locule appears bumpy with the nuclei and cells becoming less distinct. It appears that this tapetal alteration is normal during fertile pollen development (Wu and Cheng, 2000) and has been reported for the secretory tapetum in other plant systems such as in wheat (Wang *et al.*, 2003), rice (Ku *et al.*, 2003) and *Z. mays* (Warmke and Lee 1977). Accompanying this alteration, the walls of the tapetum become lined with orbicules or Ubisch bodies, which are more than just a by-product of tapetal metabolism, but contain a protein that is targeted to the microspore exine (Wang *et al.*, 2003). Given the similarities between tapetum structure and development in rice and *Z. mays,* it is possible that Ubisch bodies might also be formed on the tapetal layer in *L. perenne*.

The onset of tapetal degeneration in fertile anthers is known to be a controlled event and is marked by cell shrinkage, thinning of the cell walls, rupture of the vacuolar membrane (tonoplast), and polarization of cytoplasmic materials. This event is sometimes referred to as programmed cell death (PCD) or apoptosis of the tapetal cell layer, and has been examined in the two angiosperms *Lobivia rauschii* (a monocot) and *Tillandsia albida* (a dicot) (Papini *et al.*, 1999).

Disintegration of the tapetum results in the release of cell contents (lipids and proteinaceous molecules) that are incorporated into the developing pollen cells at the end of mitosis I and also during the later stages of pollen development (Papini *et al.*, 1999; Bedinger *et al.*, 1992).

In CMS *L. perenne*, the development of multiple small vacuoles in tapetal cells may indicate induction of a cell death programme. Premature vacuolation of tapetal cells has also been observed in CMS lines of maize. Electron microscopy further showed mitochondrial abnormalities in CMS tapetal cells at the tetrad stage (Warmke and Lee, 1977). The tapetum in the *Arabidopsis* male-sterile mutant *ms1* develops large autophagic vacuoles that are associated with necrotic cell death (rather than programmed cell death). In consequence, there was no release of tapetal cell contents into the anther locules, which may account for the abnormal wall development associated with pollen abortion in this mutant (Barrena and Wilson, 2006).

The early stages of microspore development appear to be similar in fertile and CMS *L. perenne* lines. At stage 5, fertile and sterile microspores have similarly thick exine walls, but by stage 6 the exine is noticeably thinner in sterile microspores and many have collapsed possibly as a consequence of so-called microspore leakage. Cell debris can be seen surrounding or in close proximity to the collapsed microspores in the anther locule. Evidence of microspore leakage is well-documented in the *Arabidopsis ms1* mutant, where deficient deposition of sporopollelin results in permeable exines. The microspores do not rupture but rather leak their contents which results in empty pollen sacs. This suggests that the defective *ms1* pollen wall can resist the internal pressure of the microspores, in spite of their distorted shape following cytoplasmic leakage (Barrena and Wilson, 2006).

Two kinds of sterile pollen grains were observed in *L. perenne*: either slightly irregular in shape similar to *ms1* microspores, or grains that were completely distorted and folded. The extent of distortion may reflect the condition of the pollen wall. It is possible that the completely distorted microspores are due to rupture of the pollen wall, whereas the more spherical grains may be due to leakage alone. In TGMS rice, the microspore wall is weak because it lacks primexine and sporopollenin, and in consequence, it ruptures easily as a result of pressure from the expanding vacuole, releasing cellular debris into the anther

locule (Ku *et al.*, 2003). This form of microspore leakage is more radical than the leakage reported in *Arabidopsis ms1* where bursting of microspores is not reported.

For mature pollen to be released, the anthers must undergo dehiscence. Indeed, failure to do so may on its own give rise to male sterility (Dawson *et al.*, 1993; Beals and Goldberg, 1997). In *L. perenne*, fertile anthers dehisce soon after pollen maturation whereas sterile anthers do not split at all (Connolly and Wright-Turner, 1984; Kiang *et al.*, 1993). Dehiscence depends on degeneration of the endothecium cells which are found lying between the epidermal and the tapetal layer. In sterile anthers, however, the endothecium shows no signs of degeneration, but appears to be significantly more swollen or hydrated than in fertile anthers. Degeneration is usually followed in fertile anthers by the breakdown of cells at the stomium, a structure that is contiguous with the epidermis and runs along the lateral side of each anther half (Goldberg *et al.*, 1993, Beals and Goldberg, 1997). During the dehiscence process, the endothecium becomes dehydrated, cell shrinkage occurs followed by an outward bending of the anthers, and finally with the anthers being split open (Goldberg *et al.*, 1993). In *Arabidopsis* specialized cells adjacent to the stomium undergo cell death, releasing cellular debris and calcium deposits (Beals and Goldberg, 1997). The developmentally-regulated death of these cells allows dehiscence to occur. In CMS *L. perenne* plants the lack of dehiscence might be due to abnormalities in the cell layers of the anther wall and or to the lack of mature pollen grains which results in an intra-locule pressure that is insufficient to rupture the stomium.

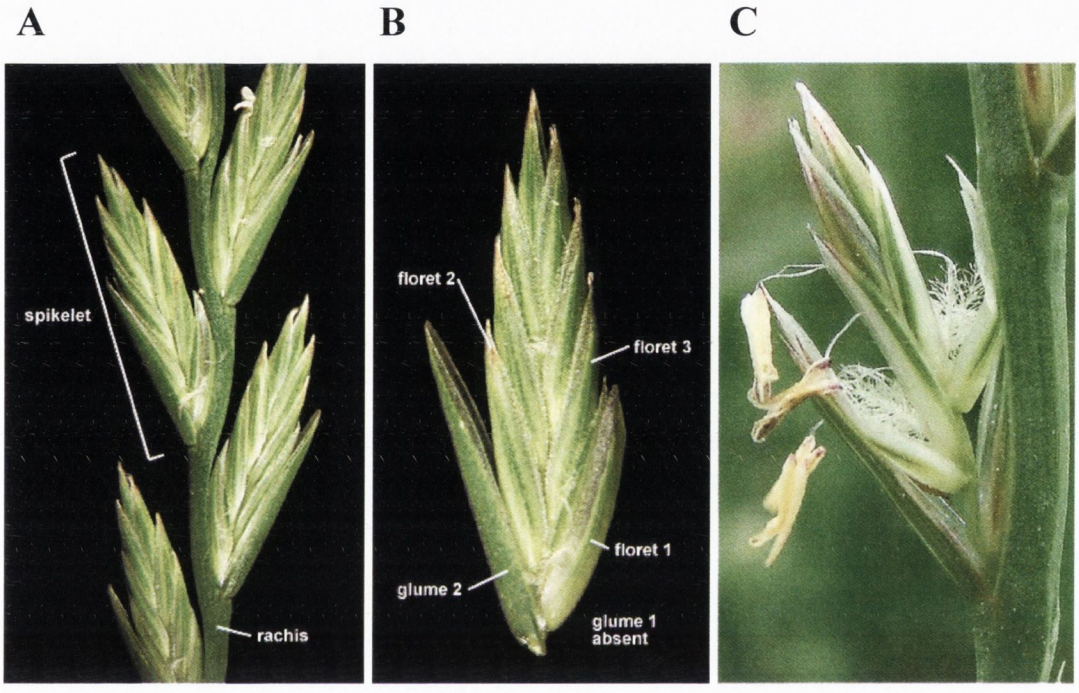


Figure 6.1 Inflorescence structure in *L. perenne*.

A. The inflorescence comprises a series of alternating spikelets on a stem (rachis). **B.** An individual spikelet. Each spikelet contains up to 9 florets at different stages of maturity. **C.** The lowest pair of florets have opened (anthesis). The fluffy white stigmas and extruded stamens are visible. The anthers have split and released their pollen.

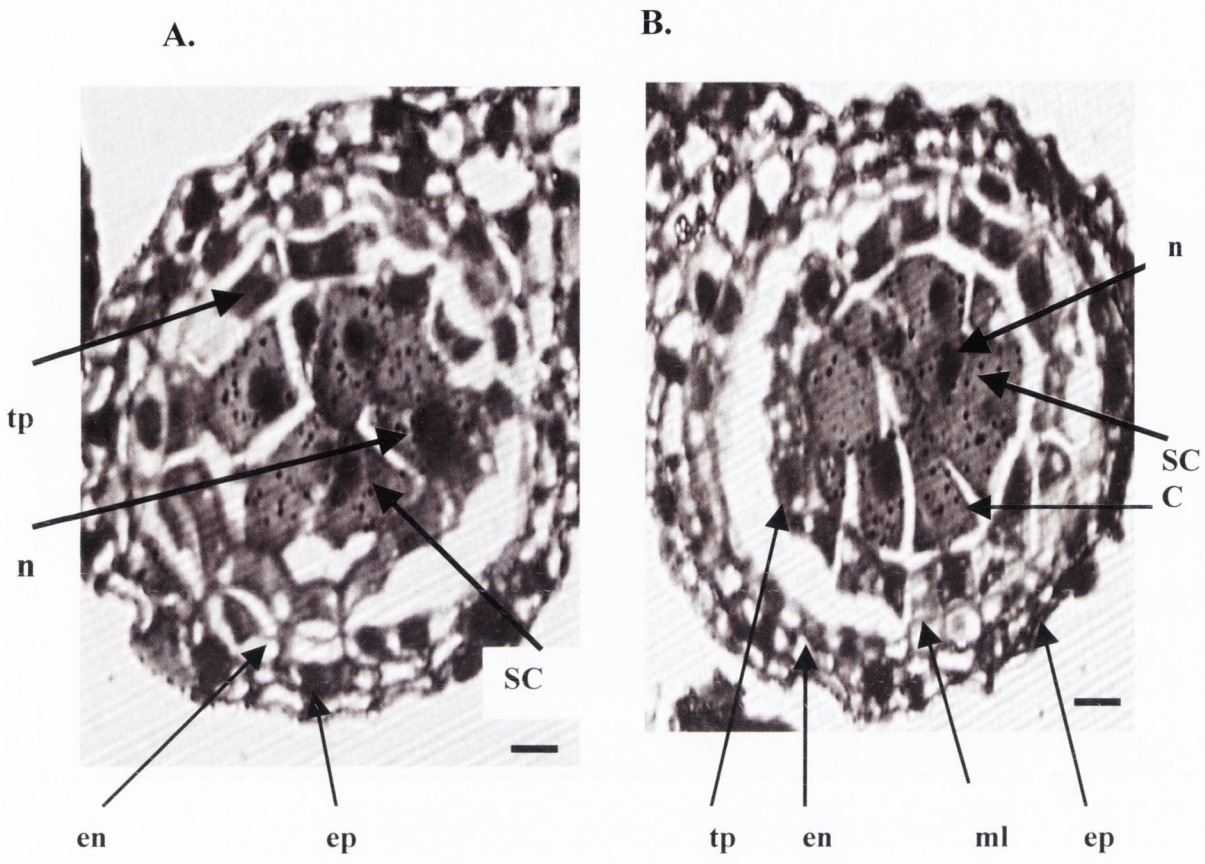


Figure 6.2 Stage ONE of pollen development in fertile and sterile line of *L. perenne*.

The earliest stage in pollen development with an un-differentiated sporogenous cell mass (SC), viewed by light microscopy (40X). Each section is 3 microns in thickness in (A) for the fertile anther and in (B) for the sterile anther. The relevant structural features of two anther cells are highlighted by arrows; Callose (C), Epidermis (ep), endothecium (en), middle layer (ml), nucleus (n) and tapetal layer (tp). Scale bar is 10 μm.

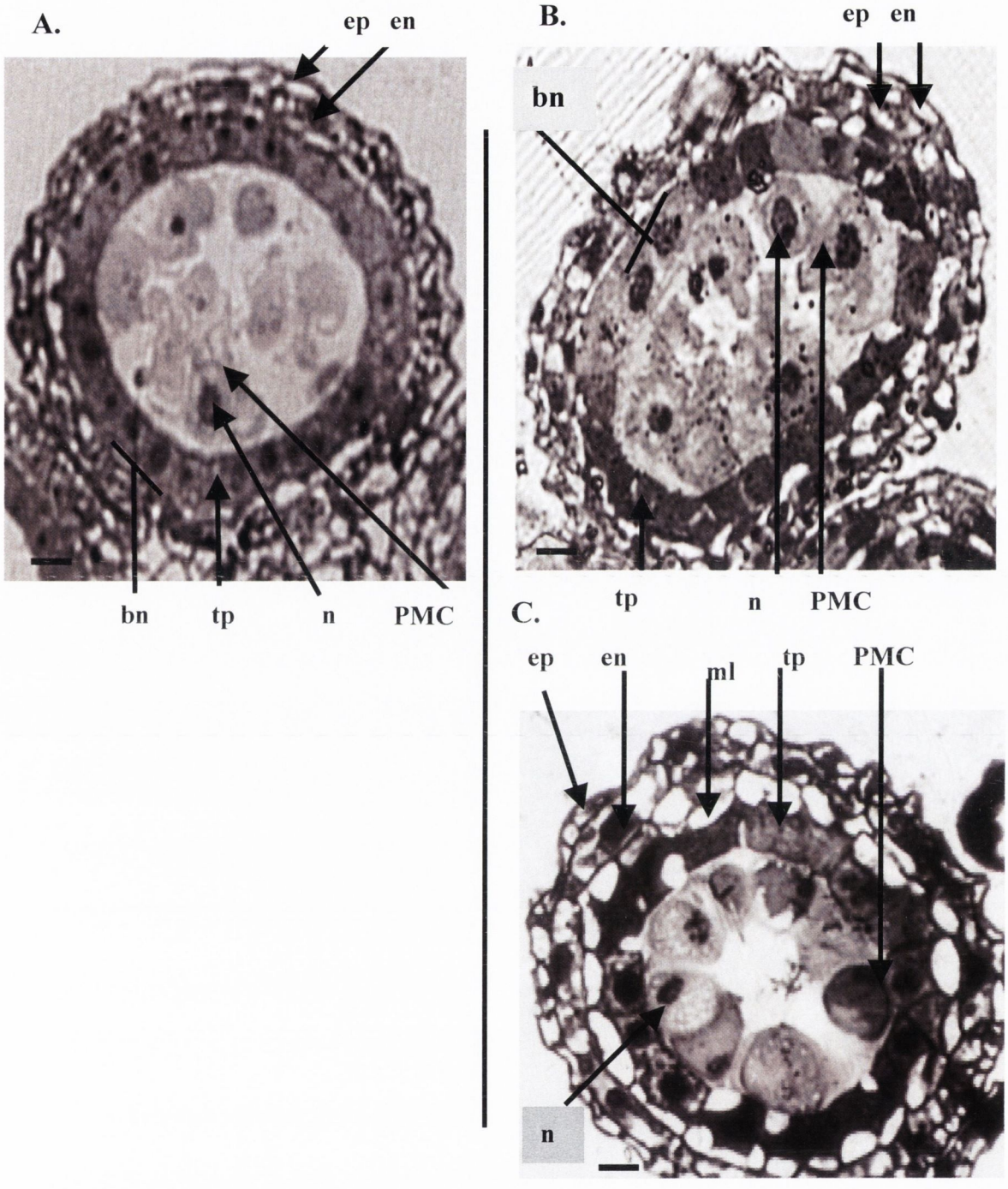


Figure 6.3 Stage TWO of pollen development in the fertile and sterile cultivar line of *L. perenne*. The second stage in pollen development with differentiated pollen mother cells (PMC). Each section is of 3 microns in thickness viewed by light microscopy (40X) in (A) for fertile anthers and in (B, C) for sterile anther sections. The relevant structural features of the comparable two anther cells in sections A, B are highlighted by arrows; Bi-nucleus (bn), Pollen mother cells (PMC), nucleus (n) and tapetal layer (tp). Also, the structural features for section C are highlighted by arrows; Bi-nucleus (bn), epidermis (ep), endothecium (en), middle layer (ml), Pollen mother cells (PMC), nucleus (n) and the tapetal layer (tp). Scale bar is 12 μm.

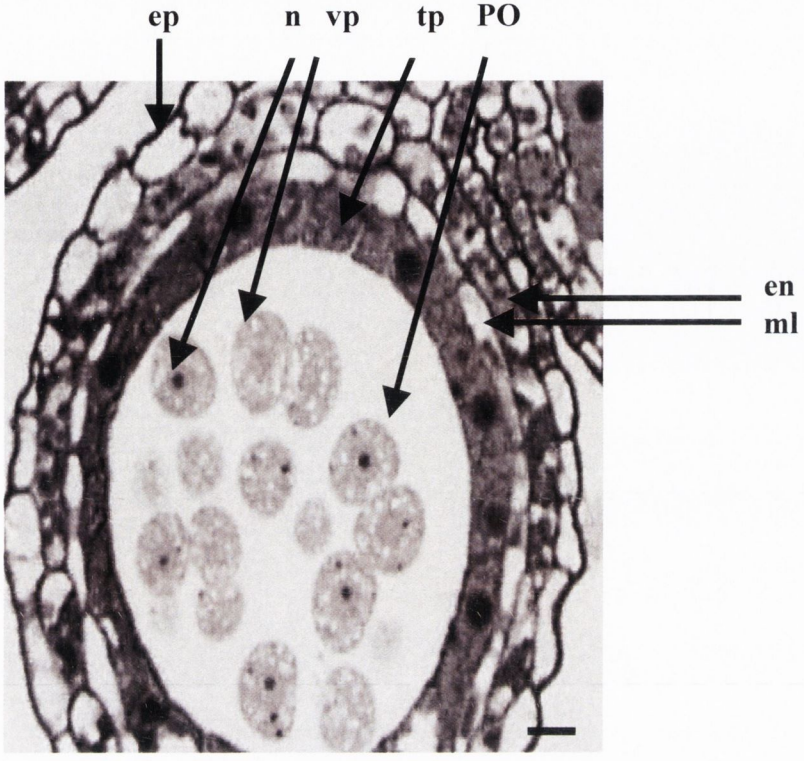


Figure 6.5 Stage FOUR of pollen development in the fertile cultivar line of *L. perenne*.

The section chosen is ca. 3 microns in thickness. The fourth stage of pollen development shows microspores dispersed in the fertile anther locule. The relevant structural features of fertile anther cells are highlighted by arrows; Epidermis (ep), endothecium (en), middle layer (ml), microspore (PO), nucleus (n), tapetal layer (tp) and vacuole microspore (vp). Scale bar is 20 μm.

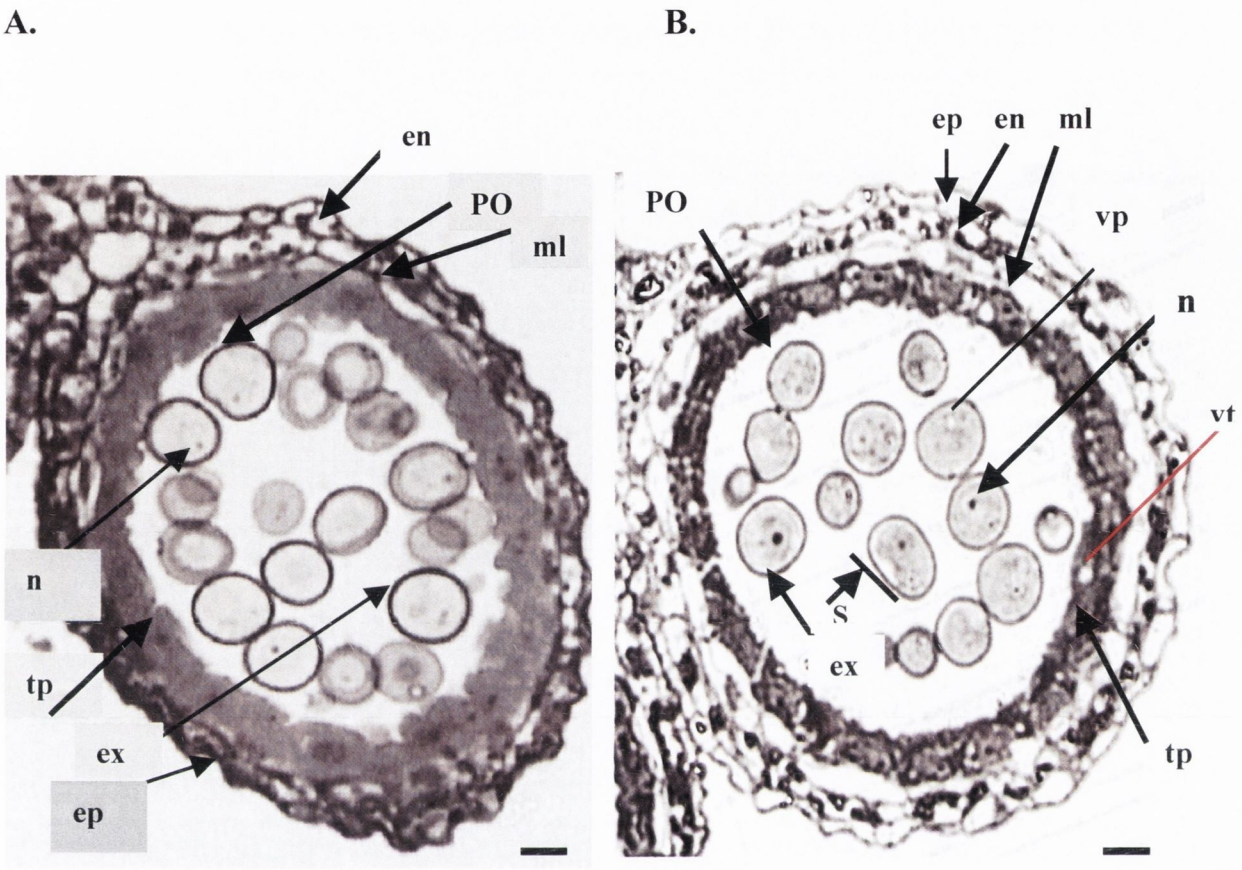


Figure 6.6 Stage FIVE of pollen development in the fertile and sterile cultivar line of *L. perenne*.

Each section is approx. 3 microns in thickness and is viewed by light microscopy (40X) in (A) for fertile anthers and in (B) for sterile anther sections. The relevant structural features of the anther cells are highlighted by arrows; Epidermis (ep), endothecium (en), exine (ex), middle layer (ml), microspore (PO), nucleus (n), tapetal layer (tp), vacuole microspore (vp), vacuole tapetal (vt), pollen pore (p), shrinkage (S). Scale bar is 35 μm.

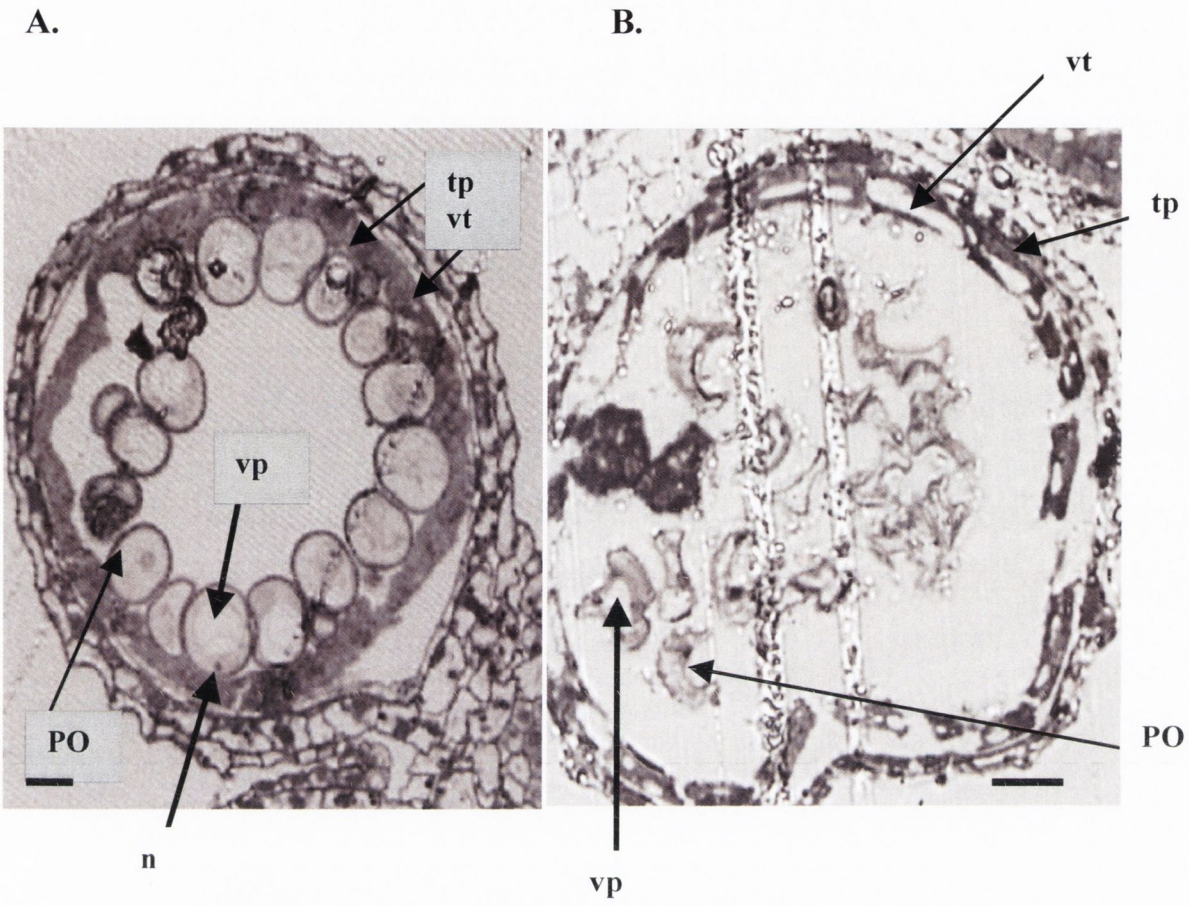


Figure 6.7 Stage SIX of pollen development in the fertile and sterile cultivar line of *L. perenne*.

The sixth stage of pollen development prior to the first mitotic division. Each section is ca. 3 microns in thickness and is viewed by light microscopy (40X magnification) in (A) for the fertile anther and in (B) for the sterile anther. The relevant structural features of two anther cells are highlighted with arrows; microspores (PO), nucleus (n), vacuole pollen (vp), vacuole tapetum (vt), tapetal layer (tp). Scale bar is 50 μm.

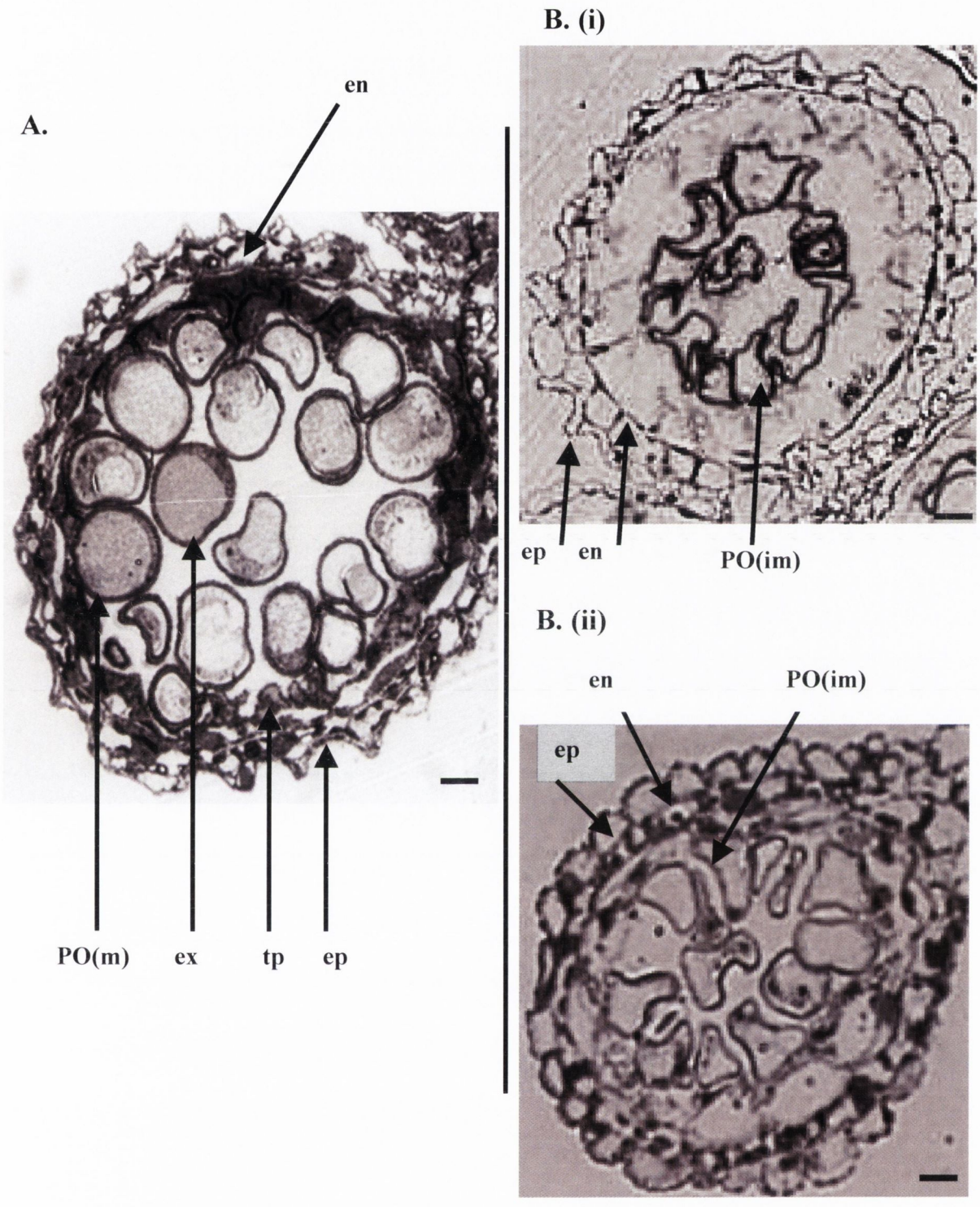


Figure 6.8 Stage SEVEN of pollen development in the fertile and sterile cultivar line of *L. perenne*. The seventh stage of pollen development during pollen maturation. Each section is ca. 3 microns in thickness viewed by light microscopy (40X magnification) in (A) for the fertile anther and in (B (i), (ii)) for the sterile anther. The relevant structural features of two anther cells are highlighted with arrows; Epidermis (ep), endothecium (en), exine (ex), mature pollen (POm), immature pollen (POim), nucleus (n), vacuole pollen (vp), vacuole tapetum (vt), tapetal layer (tp). Scale bar is 50 μm.

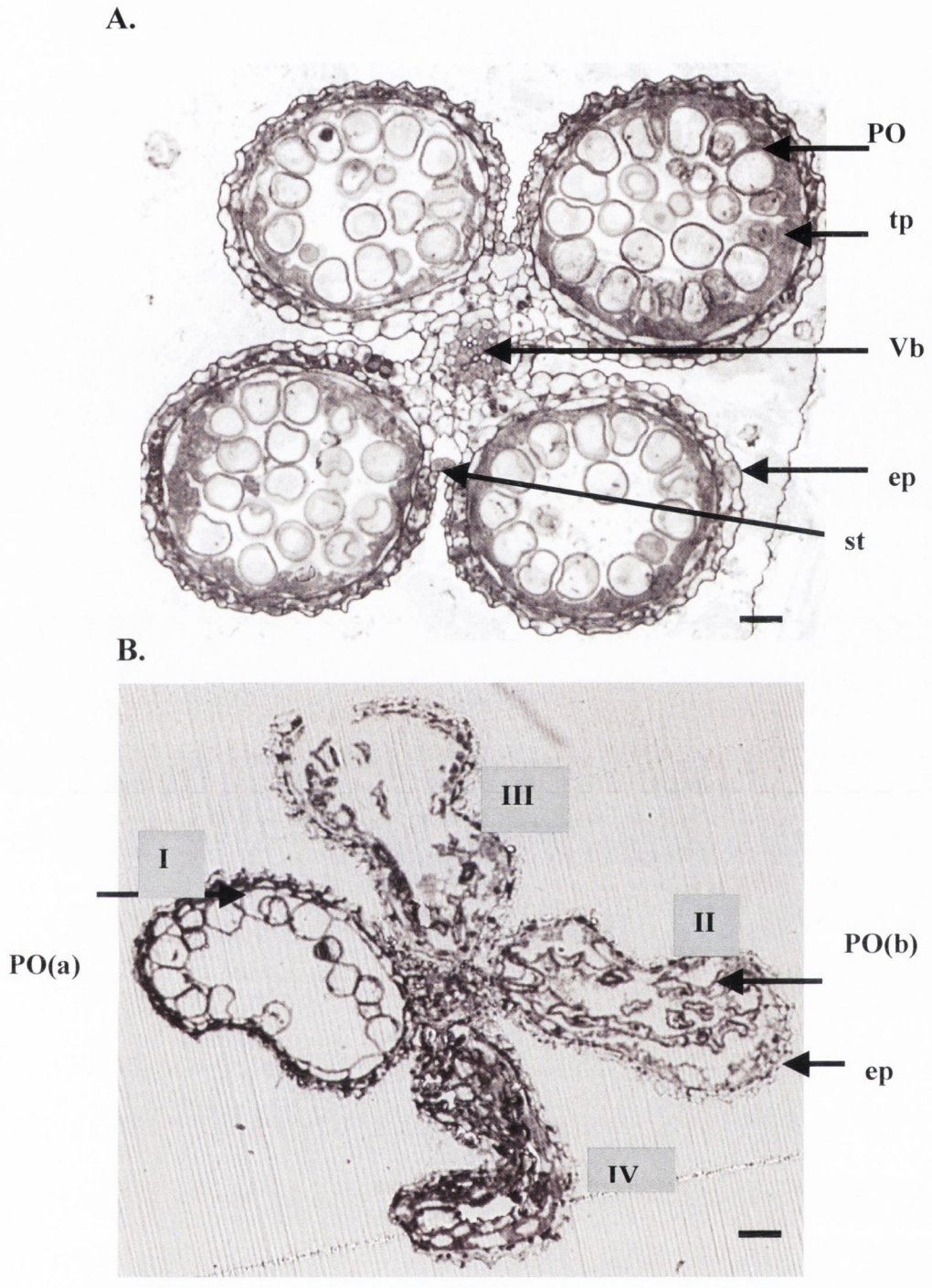


Figure 6.9 Transverse section of the anther locule from Fertile and Sterile line of *L. perenne* viewed at final stage of pollen development by Light microscopy (20X).

A. Shows a 3 micron section from the Fertile anthers containing mature pollen cells (PO), epidermis (ep), tapetum (tp), and vascular bundles (Vb). All of the four anther locules are intact. **B.** Shows a 3 micron section from the Sterile anthers which contains empty pollen cases (PO (a)), and folded pollen cells (PO (b)) in two separate anther locules (I, II), the third anther locule has burst open (III) and the fourth anther locule has collapsed (IV) and is completely folded. The epidermis (ep) is visible but anther locules are well shrunken in comparison to fertile anthers. Scale bar is 23 μm.

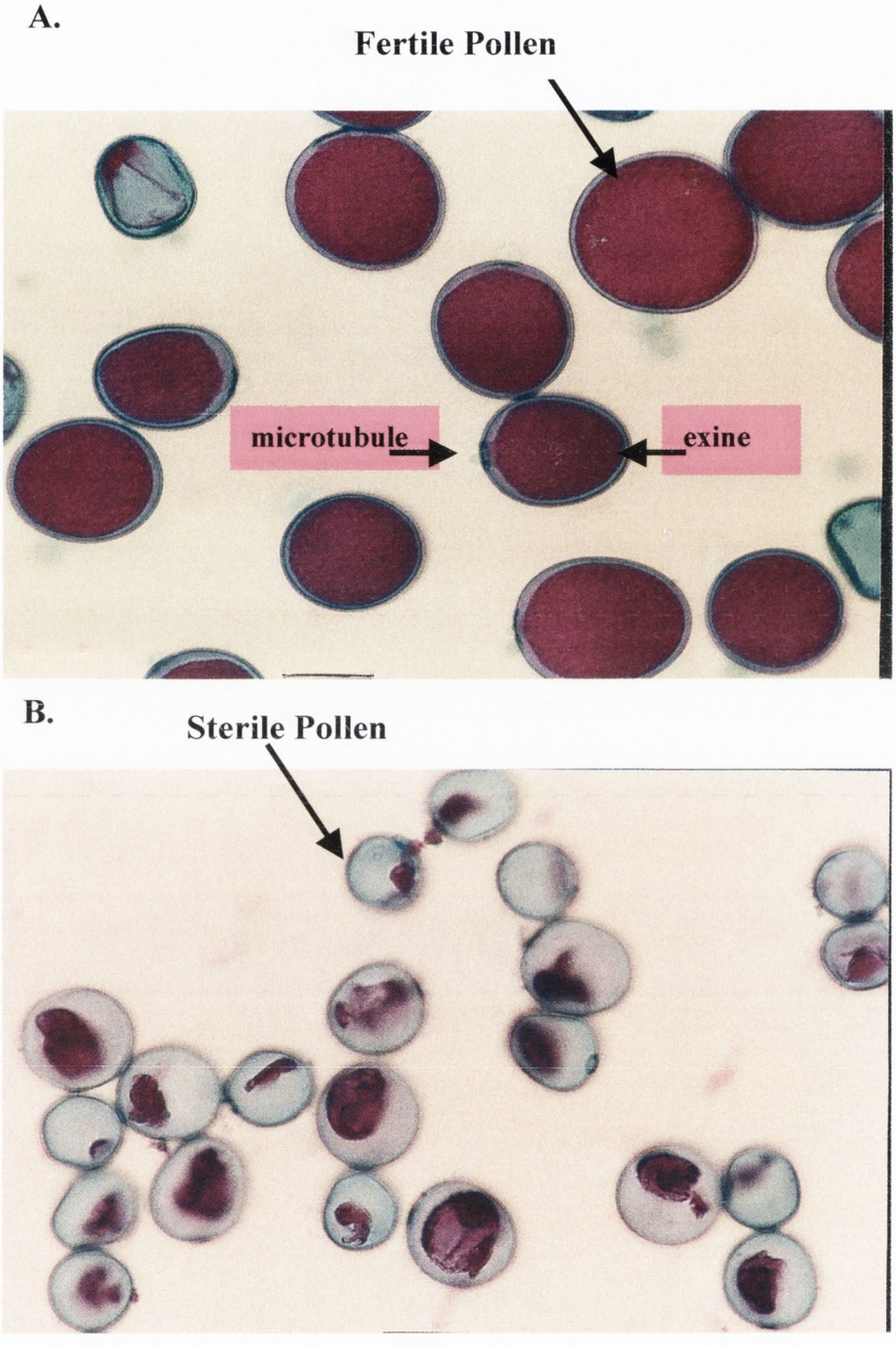


Figure 6.10 Mature pollen grains from Fertile and Sterile lines of *L. perenne* viewed by Light microscopy (20X). The pollen grains are stained with a Potassium Iodide stain (Lugol solution) to monitor for the the presence of Starch in mature pollen grains. **A.** The purple colour of the Fertile cytoplasm indicates the presence of the energy food reserve namely Starch. Also present is the outer toughened exine layer. **B.** The Sterile pollen grains are translucent or elliptical in shape, the pollen grains are smaller and the toughened outer exine coat is not as prevalent. The cytoplasm was not as intensely stained as in the Fertile line thereby illustrating a low energy reserve of starch and poorly developed pollen grains.

CHAPTER 7

Sequence analysis of *L. perenne* mitochondrial genes encoding components of the ribosome and the respiratory complexes

7.1 Introduction

7.1.1 Molecular phylogeny based on mitochondrial genes in angiosperms

In the last two decades there have been significant advances in the understanding of flowering plant evolution. The combination of information from fossil studies (Gandolfo *et al.*, 1998; Crane *et al.*, 1995) and phylogenetic studies based on nucleotide sequences has facilitated more accurate dating of key evolutionary events in the Angiosperms. According to both lines of investigation, the diversification of flowering plants is believed to have occurred during the early Cretaceous period (Crane *et al.*, 1995; Bremer, 2000). Molecular phylogeny has moreover shown that several land plant groups such as the primitive liverworts, mosses and hornworts, the Tracheophytes (vascular plants), and the Spermatophytes (seed plants) exist as monophyletic clades (Pryer *et al.*, 2002; Qiu and Palmer 1999; Knoop, 2004).

Over the past two decades organelle DNA has been regularly used in phylogenetic studies. Restriction site mapping of whole organelle DNA was initially used, for example, by Palmer and Zamir (1982). This type of analysis was mainly applied to cpDNA due to its conserved structure and size. The use of mtDNA in such studies was more problematic because it evolves more rapidly in size and structure but slowly in coding sequence (Wolfe *et al.*, 1987). With the advent of PCR, it became possible to amplify homologous fragments of coding and non-coding organelle DNA (introns or spacers) from diverse species and compare their nucleotide sequences. In both cpDNA and mtDNA, the rate of nucleotide substitutions is higher for non-coding versus coding regions (Wolfe *et al.*, 1987, 1989; Handa, 2003). Because of this trend, organelle non-coding sequences are generally used in phylogenetic analyses of closely related species

(Charmet *et al.*, 1997) whereas organelle coding regions are sometimes used in the analysis of distantly related species (Wolfe *et al.*, 1989; Laroche *et al.*, 1995).

The monocots comprise 25% of all flowering plant species and include such groups as the lilies, orchids, palms and the large grass and cereal plant family *Poaceae* which contains approximately 10,000 species. Earlier phylogenetic studies of angiosperms showed a faster rate of evolution for nuclear genes, followed by cpDNA and mtDNA which exhibits the slowest rate of evolution (10 – 20 times lower than nuclear genes). The sequence drift in coding regions is generally very low in plant mtDNAs (Wolfe *et al.*, 1987), and this can offer advantages for analysis of deeper branches in a phylogeny. However, extraordinarily large differences in the rates of mitochondrial gene evolution have been reported in certain genera. In *Pelargonium* (Parkinson *et al.*, 2005), *Plantago* (Cho *et al.*, 2004) and *Silene vulgaris* (Houliston *et al.*, 2006) mitochondrial sequence evolution occurs at a vastly accelerated rate (10-fold) relative to other angiosperm genera. The increased rate of synonymous substitutions in coding sequences in *Pelargonium* and *Plantago* has been attributed to an increase in mtDNA mutation rates or a reduction in the efficiency of DNA repair mechanisms (Parkinson *et al.*, 2005; Cho *et al.*, 2004).

A phylogeny examining the divergence of monocots from dicots, based on 9 mitochondrial protein coding genes was reported by Laroche *et al.*, (1995). Each mitochondrial gene showed variation in the synonymous nucleotide substitution rate. Averaging the substitution rate gave an estimated divergence time of 170 to 238 Myr. This was in good agreement with estimations based on chloroplast protein coding and nuclear rRNA gene substitution rates (Wolfe *et al.*, 1989) which suggested a value of 200±40 Myr. The *T. aestivum-Z. mays* and *O. sativa-Z. mays* divergence times were each estimated to be ca. 50-70 Myr based on fossil records (Stebbins GL, cited in Wolfe *et al.*, 1989). This divergence time has been used as a calibration point in studies involved in the estimation of the monocot-dicot split (Wolfe *et al.*, 1989; Laroche *et al.*, 1995) e.g. a comparison of *nad4* intron sequences between species in the *Brassicaceae* and members of the *Poaceae* (*Z. mays*, *T. aestivum*) gave an estimated monocot-dicot split of 170 to 235 Myr and a *Brassica-Arabidopsis* split as 14.4-20.4 Myr (Yang *et al.*, 1999).

Investigation of the origin of the genus *Lolium*, using cpDNA rather than mtDNA, estimated that *Lolium* diverged ca. 2 Myr ago from a common ancestor

of *Festuca pratensis*, with differentiation into the different species (which includes *L. perenne*) a more recent event, occurring ca. 1 Myr ago (Charmet *et al.*, 1997). In fact geographic studies suggest that *L. perenne* is derived from the Middle East species *L. rigidum* (Balfourier *et al.*, 1998).

7.1.2 An overview of the mitochondrial genomes of cereals

The gene content of plant mitochondrial genomes was reviewed in Chapter 1 (section 1.1). However, coding sequences (i.e. protein coding genes, ribosomal RNA and transfer RNAs) comprise a relatively small fraction of the genome e.g. 10.6% in *B. vulgaris*, 17.1% in *A. thaliana*, 9.9% in *B. napus*, 11.1% in *N. tabacum*, 6.8% in *O. sativa*, 7.9% in *Z. mays* and 7.9% in *T. aestivum.* A comparison of the mt gene content of *O. sativa*, *Arabidopsis* and *B. vulgaris* revealed differences in respect of 13 protein coding genes of which 8 encoded ribosomal proteins. Apart from known coding sequences, the remainder of the genome comprises short orfs of unknown function, most of which are species-specific, and long stretches of intergenic DNA containing many direct and inverted repeat elements. For example, in the cereals *T. aestivum*, *Z. mays* and *O. sativa,* repeated sequences account for 15.0 %, 17.3 % and 26.0 %, respectively, of the mt genome (see Chapter 1).

Mitochondrial genomes typically contain a small proportion of chloroplast sequences ranging from 1.1% in *Arabidopsis* (Unseld *et al.*, 1997) to 6.3% in O. *sativa* (Notsu *et al.*, 2002), and 2.97% in *T. aestivum* (Ogihara *et al.*, 2005) and 4.4% in *Z. mays* (Clifton *et al.*, 2004). These sequences include homologs of plastid protein coding genes, plastid tRNAs and stretches of plastid intergenic-like sequences. Remarkably, in *O. sativa,* 13.4% of the mt genome shares homology with nuclear retrotransposon and transposon-like sequences. Similar sequences were found in the mt genomes of *Z. mays* (Clifton *et al.*, 2004), *T. aestivum* (Oghihara *et al.*, 2005), *Arabidopsis* (Knoop *et al.*, 1996) and *B. vulgaris* (Kubo *et al.*, 2000).

Mitochondrial genes are often organized as clusters ranging from several genes to two-gene clusters. Clustering facilitates co-transcription from the same promoter and has been demonstrated experimentally in mitochondria of *Arabidopsis* and in *B. napus* (Brandt *et al.*, 1992; Itana and Handa , 1998). Conservation of gene clusters is poor in the cereals *Z. mays*, *T. aestivum*, *O. sativa*

(Ogihara *et al.*, 2005) with very little clustering occurring in *Z. mays*. The order of gene clusters between *O. sativa* and *Z. mays* is also poor, due to frequent gene shuffling, thereby resulting in differences between cereal mt genomes compared with the mt genomes of other plant species.

7.2 Chapter Aims

This study involved sequencing of 7 *L. perenne* mt genes encoding respiratory complex proteins, 2 encoding ribosomal proteins and the *rrn18* gene coding for the mt small subunit rRNA. The sequences were used to investigate phylogenetic relationships between *L. perenne* and other cereals including *T. aestivum* (wheat), *O. sativa* (rice), *Z. mays* (maize) and *S. bicolor* (sorghum). In addition, for the genes *atp8, cob*, *coxII*, *nad3* and *rps12*, RNA editing sites were identified by comparing the genomic sequence with that of the corresponding cDNA.

7.3 Materials and Methods

7.3.1 Materials

7.3.1.1 Plant material

Seeds of fertile *L. perenne* cv. Cashel were used in the preparation of mitochondria and mtDNA.

7.3.1.2 Primers used in the amplification and sequencing of mitochondrial genes.

The following is a list of the primers used in the amplification of 10 genes from mtDNA and 5 genes from mt cDNA templates. Forward primers are denoted by the letter F, reverse primers denoted by the letter R.

The **Complex I gene** - *nad3*, primers designed from *nad3* of *Z. mays (NB):*

nad3F 5' ATGTCGGAATTTGCACCTATTTG 3'

nad3R 5'TTACTCCCGATCCGAAGCACCC 3'

The **Complex III gene** – *cob;* primers designed from *cob* of *T. aestivum.*

cobF 5' ATGACTATAAGGAACCAACGATTCTC 3'

cobR 5' CTATTCCGTGTAATATTTTGGAATTCCTC 3'

The **Complex IV genes** – *coxI, coxIII;* primers designed from *coxI, coxIII* genes of *T. aestivum. CoxII;* primers designed from *coxII* of *L. perenne* (Kiang, S., Ph.D Thesis (1991))

coxIF 5' ATGACAAATCTGGTCCGATGGC 3'

coxIR 5' TTAGCTTTTTGTCTCTTTTACCGCAGG 3'

coxIIF 5' ATGATTCTTCGTTCATTAGAATGTCG 3'

coxIIR 5' TTAGTTGGTTTGGAGGATTAATTGATTG 3'

coxIIIF 5'ATGATTGAATCTCAGAGGCATTCTTATC 3'

coxIIIR 5' TCATGTGCCTCCCCACCAATAAATAG 3'

The **complex V genes** – *atp1, atp8,* primers designed from *atp1, atp8,* genes of *T. aestivum.*

atp1F 5' ATGGAATTCTCACCAAGAGCTG 3'

atp1R 5' TCATAAATTTAAAGCGCTTTCTTTTAAAGAAGC 3'

atp8F 5' ATGCCTCAACTTGATAAATTGACTTATTTCTC 3'

atp8R 5' TTAGATTATGCTTCCTTGCCCGTG 3'

The **ribosomal protein genes** – *rps4*, *rps12*; primers designed from *rps4*, *rps12* genes of *T. aestivum.*

rps4F 5' ATGCCTGCATTAAGATTTAAAACGTG 3'

rps4R 5' TTATATGTTTTGGCCACGTCCG 3'

rps12F 5' ATGCCTACAAAAAATCAATTGATTCGTC 3'

rps12R 5' TCATTTCGATTTAGGTCTTTCTGCAC 3'

The **small subunit rRNA** – *rrn18S*; primers designed from *rrn18S* genes of *T. aestivum.*

rrn18SF 5' ATCATAGTCAAAATCTGAGTTTGATCC 3'

rrn18SR 5' GGATTCAATCCAGCCACAGGTTC 3'

The mtDNA clones and PCR product were sequenced using a primer-walking strategy. The 5'and 3' termini of cloned PCR fragments were also sequenced using the M13 universal primers.

M13 uni (-21) 5' TGTAAAAGGACGGCCAGT 3'

M13 rev (-29) 5' CAGGAAACAGCTATGACC 3'

The following additional primers were used to sequence the mt genes referred to above. Forward primers are denoted by the letter F, reverse primers denoted by the letter R:

The **Complex III gene** – *cob*;

cobF1 5' CAGTAATTACAAGCTTAGCTAGCG 3'

cobR1 5' CCCGACCTACAAGATCCTTTAC 3'

The **Complex IV genes** – *coxI, coxII, coxIII*;

coxIF1 5' GCAATTACAATGTTATTAACCGATCG 3'

coxIF2 5' GTTAATCTGACCTTCTTTCCGATG 3'

coxIR1 5' GTGCAACCACATAATAAGTATCATG 3'

coxIR2 5' CTAAGGCTGAGCTTAATAGGAGC 3'

coxIIF1 5' GTGATTTAAGTGCAACGAAATGTAC 3'

coxIIF2 5' GAACGGTGGGAGGTCAAAGAAAGAG 3'

coxIIF3 5' GAGTAGGCGTGGAGAGCTTTTTGC 3'

coxIIR1 5' GCCTACTCTTATTTATAAAGAGCAC 3'

coxIIR2 5' GACCGGAAAGTCTGTTTGTCCACG 3'

coxIIR3 5' CTGGCATGCTAGTTCTTTTAGTCTC 3'

coxIIIF1 5' GAAATCCCTCTTCTTAATACCCC 3'

coxIIIR1 5' GAGGGATTTCCCAAGGATCTAAAAC 3'

The **Complex V genes** – *atp1, atp8.*

atp1F1 5' CATGCAAACAGGCTTAAAAGCTG 3'

atp1F2 5' GTTGAAAGCTATGAAACAAGTCTGC 3'

atp1R1 5' CAAACAGATTTGTCCATCTGTAATG 3'

atp1R2 5' GCCTTGGGAGTACCTATTGATG 3'

The **ribosomal protein genes** – *rps4, rps12;*

rps4F1 5' GATCTTTCTATATCGACATATCAGTTG 3'

rps4R1 5' CTTCTTTGTTCTTTCAAAGTCTTCTTC 3'

rps12F 5' ATGCCTACAAAAAATCAATTGATTCGTC 3'

rps12R 5' TCATTTCGATTTAGGTCTTTCTGCAC 3'

The **small subunit rRNA** – *rrn18S;*

rrn18SF1 5' GAAAGTGAAAGTCGCCAAAAAGTG 3'

rrn18SF2 5' GTAGCTGTGCTTTCAGCAAGAATTTC 3'

rrn18SF3 5' CTGTACACACCGCCCGTCACAC 3'

rrn18SR1 5' CACCTTTTTTCTTTGTTTCGGAGC 3'

rrn18SR2 5' CATGTCTCAGCAACACAAAACG 3'

rrn18SR3 5' GGTTTCGAGAGCATTCCGCCAC 3'

7.3.2 Methods

7.3.2.1 Growth of *L. perenne* seedlings

Seeds of *L. perenne* cv. Cashel were surface-sterilised in a 2% (v/v) solution of bleach (sodium hypochlorite) for a period of 20 - 30 min with gentle shaking at room temperature, washed several times in water to remove the bleach solution and were then left to soak overnight in water. After 24 hr., the rinsed seeds were thinly spread onto moistened tissue paper in deep trays. The trays were covered with aluminium foil and were placed in a darkened incubator at 28°C for a period of 5-7 days. The dark-grown seedlings were allowed to grow to a height of 4 cm (ca. 7 days growth) and were then harvested for isolation of mitochondria.

7.3.2.2 Preparation of mtDNA from seedling leaf tissue

The method used for the isolation of mitochondria from leaf tissue was described by Kiang *et al.* (1993). All steps were carried out on ice or at 4°C. Etiolated seedlings (150 g) were homogenized to a paste in 350 ml sterile Extraction Buffer (EB) (0.44 M sucrose, 50 mM Tris-HCl (pH 8.0), 3 mM EDTA (pH 8.0), 0.5% (w/v) bovine serum albumin, 10 mM beta-mercaptoethanol (added last)). A further 350 ml EB was then added, followed by further vigorous homogenization. The suspension was filtered through 4 layers of sterilised muslin and the filtrate collected in 250 ml centrifuge bottles. The filtrate was centrifuged at 1,000 x g for 10 min to pellet cell debris and nuclei. The supernatant was decanted into a fresh centrifuge bottle and was first centrifuged at 5,000 x g for 20 min to pellet chloroplasts. The recovered supernatant was then centrifuged at 16,000 x g for 20 min to pellet mitochondria.

The mitochondrial pellet was gently resuspended in 9 ml DNAase I buffer (0.44 M Sucrose, 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$) with the use of an Artist's No. 3 sterile paintbrush. DNAase I (1 ml at 10 mg/ml) was added to the mitochondrial suspension and digestion was allowed to proceed on ice for 45 - 60 min. The digestion was terminated by the addition of 0.5 M EDTA (pH 8.0) to a final concentration of 25 mM. Mitochondria were re-pelleted at 16,000 x g for 10 min. The pellet was resuspended in 20 ml of Wash Buffer (WB) (0.6 M Sucrose, 25 mM EDTA (pH 8.0), 50 mM Tris-HCl (pH 8.0). Once again the pellet was resuspended in 20 ml WB and the mitochondria re-pelleted at 16,000 x g for 10 min.

Mitochondria were lysed by resuspension in 2 ml NTE buffer (10 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)), followed by the addition of SDS to a final concentration of 1% (w/v) and incubation at 37°C for 30 min. Mitochondrial nucleic acids were purified by three phenol extractions followed by two phenol-chloroform extractions. Total nucleic acids were precipitated by the addition of two volumes of ice-cold absolute ethanol to the recovered aqueous layer. Samples were gently inverted several times and stored on ice for 60 minutes. Precipitated nucleic acids were recovered by centrifugation at 13,000 x g for 10 min. The pellet was washed with 80% ethanol, air dried and re-suspended in 200 µl sterile water. The solution was adjusted to 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0) and 3 µl DNAase-free RNAase A (10 mg/ml) was added and samples incubated at 42°C for 20 min. Extraction with phenol was used to terminate the RNAase reaction and mtDNA was re-precipitated by adding 1/10 volume 3 M sodium acetate (pH 5.2) and two volumes of cold absolute ethanol. Approximately 25 µg mtDNA was recovered from 150 g of seedlings. This was stored in sterile water at –20°C at a final concentration of 70 - 80 ng/µl.

7.3.2.3 Synthesis of cDNA from total RNA

Total RNA was extracted from fresh leaf tissue as outlined in Chapter 2. The method of single-strand cDNA synthesis from total RNA is outlined in Chapter 3.

7.3.2.4 PCR amplification from mtDNA and cDNA

PCR products of 7 respiratory complex genes (*coxI, coxII, coxIII, cob, atp1, atp8, nad3*), 2 genes encoding ribosomal proteins (*rps4, rps12*) and the *rrn18S* gene coding for 18S rRNA, were amplified from mtDNA using Platinum Taq DNA polymerase. Also, PCR products were amplified from cDNA templates for 4 respiratory complex genes (*coxII, cob, atp8, nad3*) and *rps12.* The following PCR conditions were applied: denaturation 94.0°C, annealing 60.0°C, extension 72°C for 27-28 cycles, followed by a final extension at 72°C for 5 min.

Two separately amplified PCR products for a given mitochondrial gene were individually cloned into the TOPO-TA cloning vector (Invitrogen), and each clone was bi-directionally sequenced. The sequence obtained for each mitochondrial gene (except at the 5' or 3' termini) was also confirmed by directly sequencing a single PCR product. Two PCR products for each of 5 genes were amplified from cDNA template, were individually cloned into the TOPO-TA vector, and each clone was bi-directionally sequenced. Cloned template DNAs (3 μg) and PCR products (30 ng per 100 bp) were dried down and sent for sequencing by MWG Biotech AG.

7.3.2.5 Bioinformatic resources utilised

The following databases and software were used in the retrieval and in the analysis of mtDNA sequences

► Database Analysis

NCBI	*http://www.ncbi.nlm.nih.gov/*
Blast	*http://www.ncbi.nlm.nih.gov/BLAST/*

► Sequence alignment analysis

ApE	*http://www.biology.utah.edu/jorgensen/wayned/ape*
ClustalW	*http://www.ebi.ac.uk/Tools/clustalw/*
Genejockey	*http://www.biosoft.com/biosoft*

► Phylogenetic Trees construction and evolutionary analysis

Mega align (ver 2.0 or 3.*1*) *http://www.megasoftware.net/*

7.4 Results and Discussion

7.4.1 Nucleotide sequence of mitochondrial genes of *L. perenne*

Total cellular DNA (tcDNA) and mtDNA was isolated from *L. perenne* cv Cashel and each was digested with the restriction enzymes *Hind*III, *Sal*I. In comparison with tcDNA samples, the mtDNA samples contained clearly discernible bands indicating a high level of enrichment for mtDNA (Figure 7.1). PCR products were amplified for 10 mitochondrial genes using mtDNA as the template. Primers were designed from sequences of the corresponding mt genes of either *T. aestivum,* or *Z. mays (NB).* PCR-amplified products were obtained for 7 respiratory complex genes *NADH dehydrogenase 3* (*nad3)* (Complex I); *cytochrome b* (*cob*) (Complex III); *cytochrome oxidase I – III* (*coxI*, *coxII*, *coxIII*) (Complex IV); *atp synthase 1*, *8* (*atp1*, *atp8*) (Complex V); for two ribosomal protein genes *rps4* and *rps12*, and one ribosomal RNA gene, *rrn18S*. The sizes of the amplified gene fragments, and the primer walking strategy used to sequence them are illustrated in Figure 7.2 and Figure 7.3.

The nucleotide sequences obtained for each of these genes (and that of *atp9* (Chapter 4)), were each aligned with homologous sequences from 8 higher plant species whose mitochondrial genome has been fully sequenced i.e. the four monocots *Z. mays NB* (Clifton *et al.*, 2004; Acc no.AY506529) (and *Zea mays S* (DQ490951)), *O. sativa* (Notsu *et al.*, 2002; Acc no.AB076665/6), *T. aestivum* (Ogihara *et al.*, 2005; Acc no. AP008982), *S. bicolor* (Acc no. DQ984518); and the four dicot species *A. thaliana* (Unseld *et al.*, 1997; Acc no. Y08501/2), *B. napus* (Handa, 2003; Acc. no. AP006444), *N. tabacum* (Sugiyama *et al.*, 2005; Acc. no. BA000042) and *B. vulgaris* (Kubo *et al.*, 2000; Acc. no. AP000396/7).

The extent of nucleotide sequence identity shared by the *L. perenne* genes and those of the 8 species described above is shown in Table 7.1. The identity values indicate a very high level of conservation for most of the respiratory genes, with the monocots showing a higher level of conservation in comparison with the dicots. The *coxII* sequence refers solely to the 783 bp exon and does not include the 1200 bp group II intron sequence. The ribosomal protein genes *rps4* and *rps12* also show good cross-species conservation except in the case of *A. thaliana* which has a lower than average value for *rps12*. The lowest cross-species conservation

values were obtained for Complex V genes, with *atp8* being poorly conserved in the four dicots and *atp9* showing poor conservation in some of the monocots and dicots.

7.4.2 Phylogenetic analysis of mitochondrial gene sequences

A single concatenated nucleotide sequence was prepared by joining the 11 *L. perenne* gene sequences end-to-end (i.e. the 3' end of one gene joined to the 5' end of the next) in the order shown in Table 7.1. Similarly concatenated sequences were prepared for each of the remaining 8 plant species included in the analysis. These concatenated sequences were aligned using ClustalW (Thompson *et al.*, 1994) using the default gap settings. Alignments were then checked by eye and were found to contain no abnormalities such as large gaps that might affect subsequent analyses. The resulting alignments were used to construct a phylogenetic tree of the monocot species using the dicots *A. thaliana* and *B. napus* as outgroups (Figure 7.4). Trees were constructed using a neighbour-joining algorithm (Saitou and Nei 1987) with bootstrap estimates (Felsenstein, 1985) being derived from the MEGA ver 2.1 software package (Kumar *et al.*, 2001).

The evolutionary distance between sequences was investigated by calculating K, the number of nucleotide substitutions per site as shown in Table 7.2a, and the distance matrix was used to estimate divergence dates (Myr). In this study, editing sites were not considered. Thus, evolutionary relationships were based solely on genomic mtDNA sequences. The calibration point chosen was the *Z. mays-T. aestivum* split previously estimated at 50-70 Myr (Chao *et al.*, 1984; Wolfe *et al.*, 1989). Notably, the 3 monocot species *Z. mays*, *O. sativa* and *S. bicolor* group in the same clade with a 100% BCL, while *L. perenne* and *T. aestivum* group in a separate clade, also with a 100% BCL. Therefore, the branching order for *T. aestivum* (Ta), *Z. mays* (Zm) and *O. sativa* (Os) suggests that *Z. mays* and *O. sativa* diverged more recently from a common ancestor than did either species from *T. aestivum*. This finding is in contrast to Wolfe *et al.* (1989) who based their analysis on 12 chloroplast genes which suggested that the branching order for *T. aestivum*, *Z. mays* and *O. sativa* was closer to a trichotomy, with a divergence time of 50 - 70 Myr. However they did propose that *T. aestivum* could marginally be seen to be an outgroup relative to *Z. mays* and *O. sativa*, thus

forming a dichotomy, with *Z. mays-O. sativa* forming a separate clade to *T. aestivum*. This marginal dichotomy is reflected in the branching order in Figure 7.4. Significantly, however, the *T. aestivum-L. perenne* (Lp) clade is more than marginally separated from *Z. mays-O. sativa*, thus ruling out the likelihood that the branching order is close to a trichotomy. However, a later study using a larger number of chloroplast genes contradicts our finding and those of Wolfe *et al.* (1989) by placing maize basal to a rice-wheat clade, and suggesting that maize diverged during the early stages of grass evolution from the rice-wheat lineage (Matsuoka *et al.*, 2002). As limited data is available regarding grass phylogenetic studies based on organelle sequences, there appears to be inconsistent findings that depend to some extent on the organelle or on the particular genes chosen. Thus, an improved phylogeny of the cereals would ideally require evaluation of a greater number of genes from both plastids and mitochondria.

Confidence in the phylogeny derived from the concatenated nucleotide sequences was supported by re-analysing phylogenetic relationships using concatenated amino acid sequences derived from the 10 protein-coding genes (*rrn18* was excluded form this analysis) (Figure 7.5). These alignments were also checked by eye for abnormalities such as large gaps but none were found. An almost identical tree with similarly high BCL values was obtained in which *Z. mays*, *O. sativa* and *S. bicolor* are clustered with an 89% BCL value. *L. perenne* and *T. aestivum* again cluster together with a BCL of 100 %, indicating their possible common ancestry.

The distance values calculated for each pair of nucleotide and amino acid sequences (Table 7.2a and b, respectively) was used to determine the evolutionary relationships. Incorporating the previously estimated *T. aestivum-Z. mays* divergence time of 50 - 70 Myr as a calibration point (Wolfe *et al.*, 1989;Yang *et al.*, 1999) allowed us to estimate divergence dates for the various species. The number of substitutions per site (K) between *T. aestivum* and *Z. mays* is 0.01784 (Table 7.2a). The K values for other species pairs allowed us to generate an estimated divergence date for the selected monocot species (Table 7.3), which placed the *L. perenne-T. aestivum* split at between 34.2 and 47.8 Myr. The *O. sativa-Z. mays* divergence time is therefore 36.4 - 51.0 Myr, which is younger than the earlier estimation of 50 - 70 Myr predicted from fossil records (Stebbins, 1981 cited by Wolfe *et al.*, 1989). All of these monocot species belong to the

grass family *Poaceae*, and studies examining the lineages of monocot flowering plants using the chloroplast *rbc*L gene placed the origin of this family at more than 70 Myr (Bremer, 2000). The divergence time between *Brassica-Arabidopsis* in our study is estimated at 28.4-39.7 Myr which is twice as old as the *Brassica-Arabidopsis* split estimated from a single mitochondrial gene by previous workers (Yang *et al.*, 1999).

Our study gave an estimated monocot–dicot divergence time of 218.6 – 306.1 Myr which was calculated by averaging the distances between the six monocot and the two dicot species (Table 7.2a, Table 7.3). Earlier studies estimated the monocot-dicot split from plastid genes and nuclear rRNA genes to be 160 to 200 Myr (Wolfe *et al.*, 1989). This estimate is in close agreement with the estimated divergence time of 170 – 235 Myr by Yang *et al.* (1999) who based their analysis on mitochondrial intron sequences of 10 species in the *Brassicaceae* and the two monocots *Z. mays* and *T. aestivum*. Other workers examined synonymous substitution rates separately in 9 mitochondrial genes and in 5 mitochondrial genes using the *Z. mays-T. aestivum* split of 50-70 Myr as the landmark calibration time, giving an estimated monocot-dicot divergence time of 170-238 Myr and 157-226 Myr, respectively (Laroche *et al.*, 1995). The monocot-dicot split estimated from our data is older, at 218 to 306 Myr. Finally, we obtained a nucleotide substitution rate of 0.127 - 0.178 x 10^{-9} substitutions per site per year for the *Z. mays-T. aestivum* split by dividing the nucleotide substitution rate (K = 0.01218) by twice the estimated divergence time (T) of 50 - 70 Myr (Chao *et al.*, 1984; Wolfe *et al.*, 1989), which is based on the equation K/(2T) (Li and Graur, 1991; Kimura, 1980). This value is slightly lower than the substitution rate of 0.16 – 0.23 x 10^{-9} in *Z. mays-T. aestivum* calculated for the mitochondrial *nad4* intron (Yang *et al.*, 1999). All of these calculations must, however, be treated with caution as the estimates are based on limited sequence data and essentially give only a rough guide to the evolutionary history of the genera involved.

The Kimura method (Kimura, 1980) facilitated comparison of the number of nucleotide substitutions per site for the mt genes across several genera, from which estimates of their divergence times could be estimated, but it did not distinguish between synonymous and non-synonymous substitutions. Therefore, in conjunction with divergence time estimates based on nucleotide substitutions

per site, the number of synonymous substitutions per site was also calculated for the various genes by the method of Li-Wu-Luo (Li *et al.*, 1985) using the MEGA software. The nine genes chosen and the number of synonymous substitutions per site (Ks) are listed in Table 7.4a. The genes *atp9* and *nad3* were not used in this analysis for the following reasons: (i) the number of synonymous substitutions per site (Ks) between *T. aestivum* and *L. perenne* for *nad3* was not statistically different than zero; (ii) although two *L. perenne atp9* gene sequences were available (from *L. perenne* CMS9B290, LPS21) the sequence of the Cashel cultivar *atp9* gene was not. Also the average number of synonymous substitutions per site for *atp9* between *L. perenne* and the other monocots, Ks = 0.35452 (data not shown), was statistically greater than the average number of synonymous substitutions per site between *L. perenne* and the dicots, Ks = 0.32024 (data not shown).

The ratio value in Table 7.4b is the number of the synonymous substitutions per site for *L. perenne-T. aestivum* (Ks = 0.0255) relative to *Z. mays-T. aestivum* (Ks = 0.03950) (Table 7.4 (a)) which was multiplied by the estimated divergence time of 50 – 70 Myr for the *Z. mays -T. aestivum* split (Wolfe *et al.*, 1987, Laroche *et al.*, 1995). This gave a divergence estimate of 32.2 – 45.1 Myr for the *L. perenne-T. aestivum* split and 35.9 – 50.3 Myr for the *Z. mays-O. sativa* split. The monocot-dicot divergence in our study (Table 7.4 (b)) is placed at 228.2 – 319.5 Myr, being older than the estimated divergence time of 160 - 240 Myr (Wolfe *et al.*, 1989). Although there is close agreement for the time estimates between the two systems, one is making an assumption of rate homogeneity i.e. that the various organelle gene sequences evolved at a constant rate. However, according to several studies there can be considerable rate heterogeneity between genes in the same genome as well as between genes in different genome compartments, and between species (Wolfe *et al.*, 1987; Chaw *et al.*, 1997). In the earlier studies (Martin *et al.*, 1989; Wolfe *et al.*, 1989; Brandt *et al.*, 1992) a statistical assessment for homogeneity in the substitution rates was not considered. A later study (Laroche *et al.*, 1995) included only sets of gene sequences showing statistically homogenous rates of synonymous substitution. In our study (Table 7.1) the majority of the mitochondrial genes selected (*coxI*, *coxII*, *coxIII*, *atp*1, *rps*12, *cob*) showed homogenous substitution rates across the chosen taxa in the Laroche *et al.*, (1995) study. Although rate homogeneity was

not evaluated for the remaining genes (*rps4*, *nad3* and *rrn18S*) it was assumed that they were also homogenous.

In my study when the average number of synonymous substitutions per site (Ks) was divided by twice the divergence time (T) between *Z. mays* and *T. aestivum* (Ks x (10^{-9})/2T), the synonymous substitution rate per year was calculated as 0.28 - 0.39x10^{-9}. This value is lower than the value of 0.4 - 0.56 x 10^{-9} calculated for several mitochondrial genes in relation to the *Z. mays-T. aestivum* split by Laroche *et al.* (1995), but is slightly higher than for the 2 mitochondrial genes used by Wolfe *et al.* (1989). Similarly, the overall number of synonymous substitutions per site obtained for the monocot-dicot split in my study gives a value of 0.18 (Table 7.4a) which is similar to the estimates of Laroche *et al.* (1995) who reported an overall value of 0.19 for 9 mitochondrial genes.

Given the lack of full mitochondrial genome sequences for *Lolium* and *Festuca*, my study focused solely on the phylogenetic relationships of *L. perenne* with distantly related genera of the *Gramineae* family. Earlier studies (Charmet *et al.*, 1997) using cpDNA and nuclear ITS (internal transcribed spacer) sequences generated a phylogenetic relationship for *Lolium/Festuca* and other *Poaceae* species. The *Festuceae* diverged from *Poa trivialis* ca 13 Myr ago and split into broad-leaved and fine-leaved fescues around 9 Myr ago. In the broad-leaf group the genus *Lolium* diverged about 2 Myr followed by differentiation into distinct species ca 1 Myr ago.

7.4.3 RNA editing in *L. perenne* mitochondrial genes

RNA editing was assessed for 6 *L. perenne* genes: *coxII*, *atp8*, *cob*, *rps12*, *nad3* (Figure 7.6) and *atp9* (Chapter 3). The primary concern was the identification of full editing sites thus identifying synonymous and non-synonymous codon substitutions. The presence of partial editing sites identified, for example, in *B. vulgaris* (Mower and Palmer, 2006), was not considered in our study because only two cDNA clones per gene were sequenced and direct sequencing of RT-PCR products was not considered. In total, 59 editing sites were found in four genes i.e. *coxII, atp8, cob* and *nad3;* no editing sites were detected in *atp9* (Chapter 3) and *rps12*. All of the editing sites were C to U conversions except for a single U to C conversion in *nad3*. No new start (ATG) or

stop codons (TAA, TGA, TAG) were created by the editing process for any of the genes investigated.

The percentage nucleotide identity between the six *L. perenne* cDNAs and homologues from other plant species is shown in Table 7.5. The average percentage identity for edited genes was highest for *O. sativa* at 97.73%, being marginally greater than the value for *T. aestivum* at 97.39% followed by *S. bicolor*, *Z. mays(NB)*, *A. thaliana, B. napus* and *B. vulgaris* in descending order (Table 7.5A). This is in marked contrast to comparisons of the corresponding genomic DNA sequences (Table 7.1). The nucleotide identity for *T. aestivum* at 97.57% is marginally higher than *O. sativa* at 97.55% followed by *Z. mays (NB)*, *S. bicolor, B. vulgaris, B. napus* and *A. thaliana* in descending order (Table 7.1). In Table 7.5B examination of percentage identity values for the predicted amino acid sequences derived from cDNA for each of the above plant species, relative to *L. perenne,* follows a pattern similar to the cDNA sequence comparisons (Table 7.5A) with the highest identity value occurring in the monocots for *O. sativa* and *T. aestivum,* followed by *S. bicolor*, and *Z. mays(NB)*.

7.4.3.1 Variation in RNA editing frequency and in the abundance of silent and non-silent editing sites in mt genes of *L. perenne* and other monocot species

In *L. perenne,* the frequency of editing varied across the six genes investigated: 0% for both *atp9* and *rps12*, 0.8% for *atp8*, 1.63% for *cob*, 2.3% for *coxII* (exon), and 5.04% for *nad3*. Variation in editing frequency as a function of gene size has also been detected for the same six mt genes in the monocots *T. aestivum*, *Z. mays* (NB), *O. sativa* and *S. bicolor,* and in the dicots *A. thaliana, B. napus, B. vulgaris* (Table 7.8). In *L. perenne*, for example, the 357 bp *nad3* gene contains 18 edited sites, *cob* at 1.2 kb contains 19 edited sites, while *rps12* (378 bp) contains none. The editing frequency of a given gene across different species, including *L. perenne,* is reasonably consistent, implying that genes that are frequently edited in *L. perenne* are also frequently edited in other higher plant species (Table 7.8). In addition, *L. perenne* contained a larger proportion of non-silently edited sites (i.e. codon-changing editing events) than silently edited sites. In total, 59 sites were edited affecting 55 codons, in cDNA sequence totalling 3,379 bp. Only 7 out of 55 codons were silently edited (Table 7.6). Non-silent

editing gave rise to 3 amino acid alterations in *atp8*, 13 in *nad3*, 15 in *coxII* and 17 in *cob* (Table 7.6). Of the silently edited sites one occurred in *atp8*, and two each in *coxII, cob, nad3* (Table 7.7).

The *coxII* gene (2.067 kb) consists of a 783 bp exon and a 1.284 kb group II intron (Figure 7.6C). Our analyses focused solely on the exon which is frequently edited in *L. perenne, Z. mays (NB), O. sativa, T. aestivum, A. thaliana* and *B. napus*. The *L. perenne* genomic sequence is very similar to that of other members of the *Gramineae* with values of 99.6%, 99.2%, 99.1%, 99.0% for *T. aestivum*, *O. sativa*, *Z. mays* and *S. bicolor* respectively, in contrast to the dicot species which range from 90.9% to 92.3% (Table 7.1). A comparison of RNA editing of *L. perenne coxII* with editing in the above graminaceous species gives a different series of identity i.e. 99.1%, 98.72%, 98.59% for *O. sativa*, *Z. mays* and *T. aestivum*, respectively (Table 7.5A). Thus, following RNA editing, the identity values for the *coxII* exon slightly decrease for all three species with *O. sativa* showing the greatest similarity followed by *Z. mays* and *T. aestivum*. The *L. perenne coxII* exon contains 18 editing sites, two of which are silent edits i.e. the codons CTT (Leu-11) and CTT (Leu-48) (Table 7.7, Figure 7.7). Both of these are conserved in the amino acid sequence (predicted from cDNA) across three of the monocots (Figure 7.7, Table 7.7). *O. sativa* contains 19 editing sites, of which 17 are identically positioned with *L. perenne*, amounting to 89.9 % editing site identity (Table 7.8), whereas *T. aestivum* gave an editing site identity of 87.5 %, with 14 of its 16 sites the same as in *L. perenne* (Table 7.7, Figure 7.7A). The amino acid substitutions that occur as a result of *coxII* editing in *L. perenne* are highly conserved across the three monocot species (Figure 7.7), thus emphasising the importance of these residues in coxII protein function. An amino acid identity value of 99.6 % between *O. sativa* and *L. perenne* was the highest among the monocots (Table 7.5B, Figure 7.7B). Only the amino acid residue at codon 5 differed between these two species, due to a C to T transition in *O. sativa* (Figure 7.7), changing a TCA (Ser-5) to a TTA (Leu-5) codon in both *O. sativa*, and *Z. mays*; editing does not, however, occur at this site in *L. perenne* or in *T. aestivum* (Figure 7.7). Likewise the codons TTA (Leu-11), ACT (Thr-130) which are not edited in *L. perenne* or *O. sativa*, were edited in *Z. mays* and *T. aestivum*, respectively (Figure 7.7).

The *atp8* gene (Figure 7.6), formerly known as *orfB* (Sabar *et al.*, 2003; Unseld *et al.*, 1997) is present in plant, but absent in mammalian mt genomes. Editing of *atp8* appears to have been examined only in plant species *O. sativa*, *B. napus* and *B. vulgaris*, and as in *L. perenne*, it occurs at a low frequency (Table 7.8). Most notably the editing frequency and the position of edited sites was identical for *L. perenne* and *O. sativa* (Table 7.8, Figure 7.8A), and any differences in nucleotide sequence between these two species was due to point mutations in the genomic sequences and not to editing. For this reason, the amino acid identity between these two species was 94.84%, (Table 7.5B, Figure 7.8 A, B). The single silently edited CTT (Leu-41) codon in *L. perenne* (Table 7.6) is conserved in the cDNA coding sequence of *O. sativa* (Figure 7.8). In fact, all four editing sites between these two species are identified as being silent (Figure 7.8A).

Editing of the *cob* (Figure 7.6A) gene has not been determined in the majority of monocot species sequenced to date, but as reported in Genbank, 19 sites are edited in *O. sativa* which is the same as the number of sites edited in *L. perenne* (Table 7.6, Table 7.8, Figure 7.9A). Of these, 16 C to T sites are identically positioned, thus giving an editing identity value of 84.2%. The two silently edited codons CTT (Leu-188), TTT (Phe-352) in *L. perenne* (Table 7.6) are conserved in *O. sativa cob* cDNA (Figure 7.9, Table 7.6). Due to the high editing site similarity, there is complete conservation of the amino acid encoded at each edited sites in the two species. However, two amino acid differences, not due to editing are located at the carboxy terminus: TTC (Phe-363) and GAT (Asp-388) in *O. sativa* (Figure 7.9B), resulting in an overall amino acid identity of 99.48% (Table 7.5B).

Transcripts of *rps12* (Figure 7.6B) are not edited in *L. perenne* or *O. sativa* (Table 7.8). Interestingly, however, non-silent editing occurs at codon positions 23, 66, 74, 90, 95 and 97 (Figure 7.10) in *T. aestivum*, *Z. mays* and *S. bicolor* (Figure 7.10 A, B). A single amino acid difference, not due to editing, exists between *L. perenne* and *O. sativa* because of a base change in codon 16: CAG (Gln-16) in *O. sativa* is replaced by CGG (Arg-16) in *L. perenne,* and in *Z. mays* (NB), *S. bicolor* and *T. aestivum.* Previous workers identified a heterogenous population of *rps12* cDNAs in *T. aestivum* (Gualberto *et al.*, 1991), *Z. mays* (Phreaner *et al.*, 1996) and *Petunia* (Lu *et al.*, 1996), indicating the presence of

completely and incompletely edited sites. Since a nuclear copy of the *rps12* gene has not been detected in *O. sativa* (Notsu *et al.*, 2002), this implies that mt *rps12* is functional, despite lacking the conserved amino acid residues (Figure 7.10). This may indicate that the retention of *rps12* in the mt genome of *Lolium* also implies functionality. We do not yet know if a nuclear copy of *rps12* exists in *L. perenne*. In general, over the course of evolution, plant ribosomal protein genes tend to be lost to the nucleus, thus indicating a bias with respect to the translocation of genes involved in organelle ribosome function (reviewed in Palmer *et al.*, 2000).

In *L. perenne, nad3* (Figure 7.6B) is edited at 18 sites which is consistent across the monocots (Table 7.8). A large number of these sites are identically positioned in the *nad3* genes in other monocots (Table 7.8, Figure 7.11). For example, in *T. aestivum* 16 of its 21 edited sites are positioned identically with *L. perenne* (Table 7.8, Figure 7.11A) and 19 represent silent alterations between these two species (Figure 7.11A). Silent editing at codon positions 27, 49, 64, 70, 83, 84, 92 is conserved across *L. perenne, Z. mays, T. aestivum, O. sativa, S. bicolor* (Figure 7.11). In addition, the silently edited codons ATT (Ile-13) and GGC (Gly-50) in *L. perenne* (Table 7.6), are both conserved in the amino acid sequence in each of the chosen monocots (Figure 7.11B). However, two additional conserved editing sites occur in these species that do not occur in *L. perenne* i.e. at codon positions TCG (Ser-115) and CGG (Arg-117) (Figure 7.11). In total there are 8 amino acid differences, across the four different monocot species (highlighted in red type in Figure 7.11B), seven of which are due to the editing process. Although not investigated in *L. perenne*, the *nad3* gene is found to be co-transcribed with *rps12* in *T. aestivum* (Gualberto *et al.*, 1991), *Petunia* (Lu *et al.*, 1996) and *Oenothera* (Schuster *et al.*, 1991) with RNA editing being controlled by the same nuclear locus for both genes.

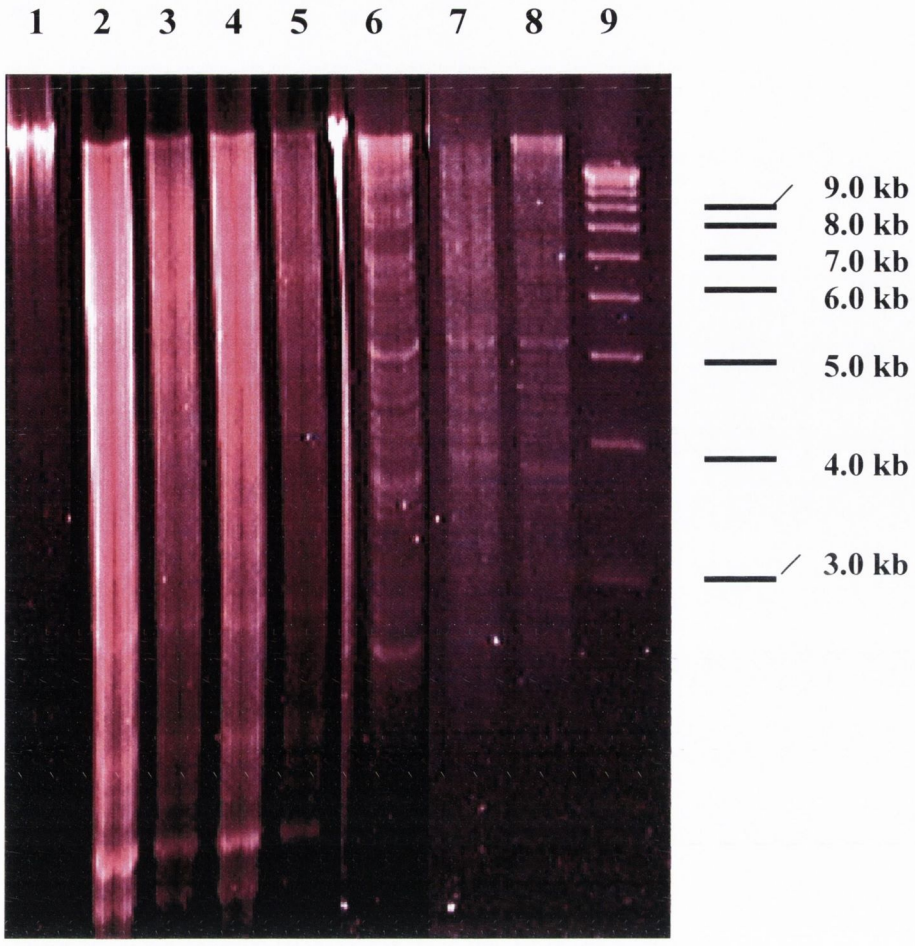
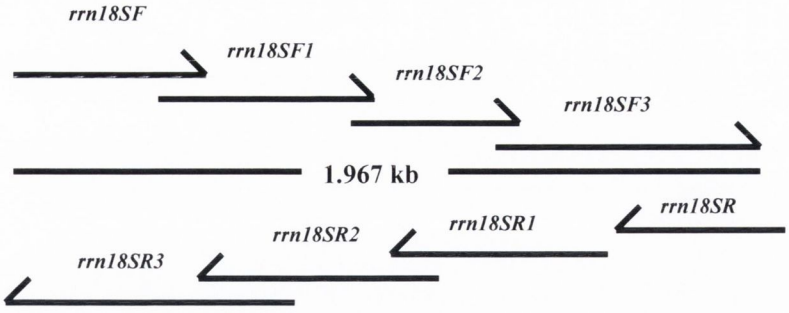


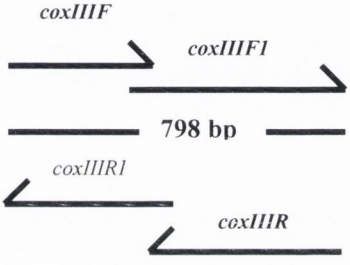
Figure 7.1 Restriction digestion fragments of mitochondrial (mt) DNA and total cellular (tc) DNA in *L. perenne*.

Lane 1 is undigested mtDNA. lane 2 *Hind*III digested tcDNA in the *L. perenne* lines CMS9B290, lane 3 *Hind*III digested tcDNA in the *L. perenne* maintainer line LPS21 respectively, lanes 4 and 5 *Hind*III digested tcDNA in *L. perenne* cultivar line Cashel (L). Lanes 6 and 8 *Hind*III digested mtDNA in *L. perenne* cultivar line Cashel (L), lane 7 *Sal*I digested mtDNA in *L. perenne* cultivar line Cashel (L), lane 9 is the 1 kb ladder.

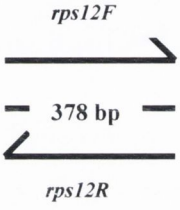
(i) **rrn18S**



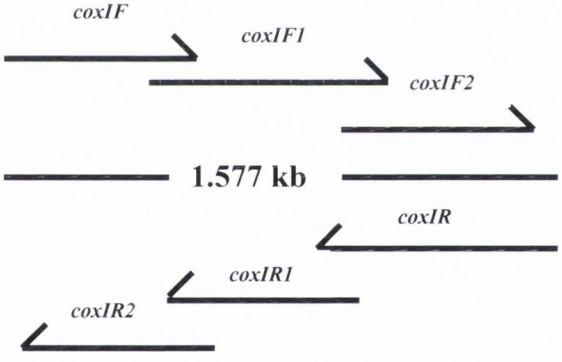
(ii) **coxIII**



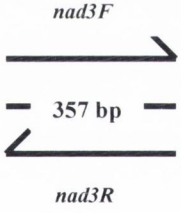
(iii) **rps12**



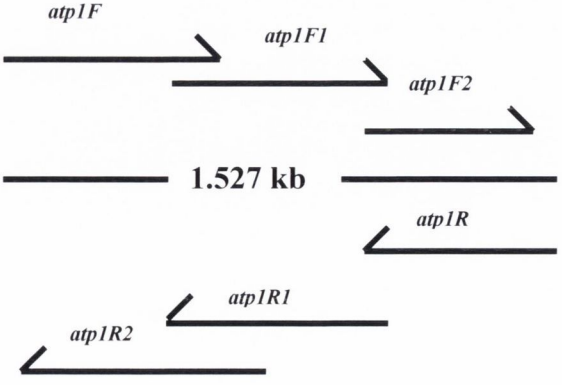
(iv) **coxI**



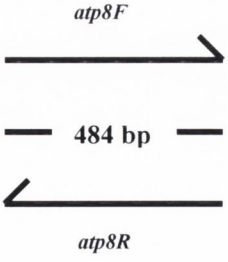
(v) **nad3**



(vi) **atp1**



(vii) **atp8**



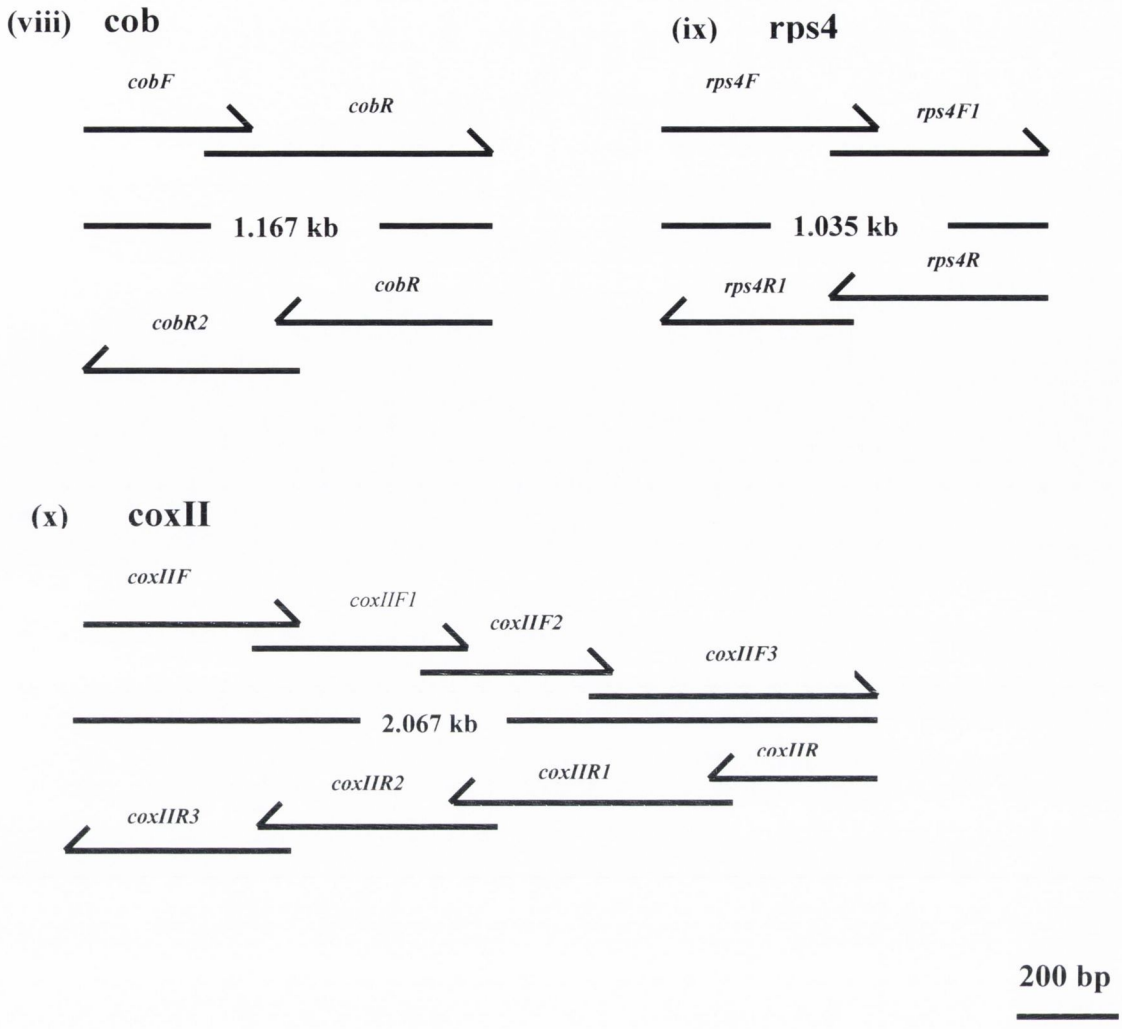
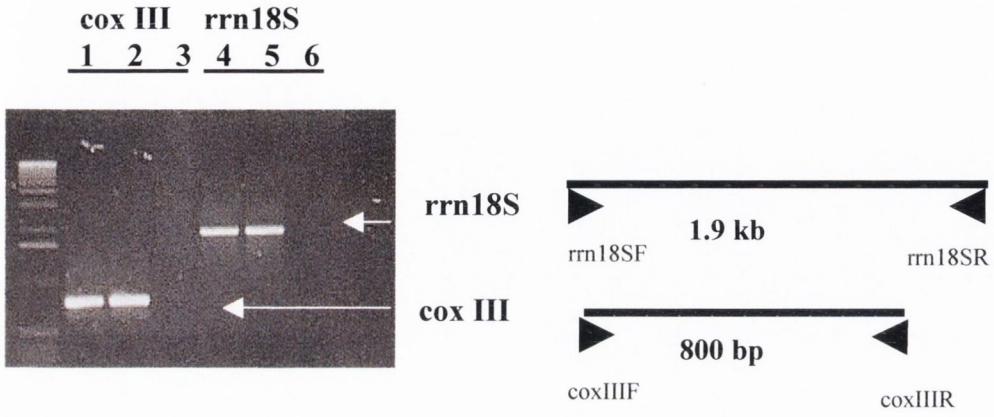


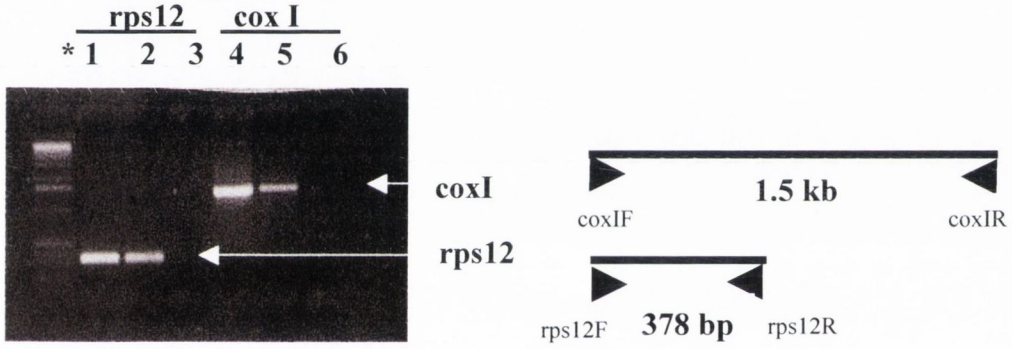
Figure 7.2 Position of primers used in the bi-directional sequencing of the clones for 9 mitochondrial genes and a rRNA in *L. perenne*.

The mitochondrial genes sequenced are; (i) *coxIII* (ii) *rps12* (iii) *coxI* (iv) *nad3* (v) *atp1*(vi) *atp8* (vii) *cob* (viii) *rps4* (ix) *coxII* and the rRNA is *rrn18S*. Each of the arrows shown, highlights the size and the direction sequenced by primer walking strategy. The size for each of the genes sequenced is given in base pairs (kb/bp). Name of primer used is indicated above each arrow.

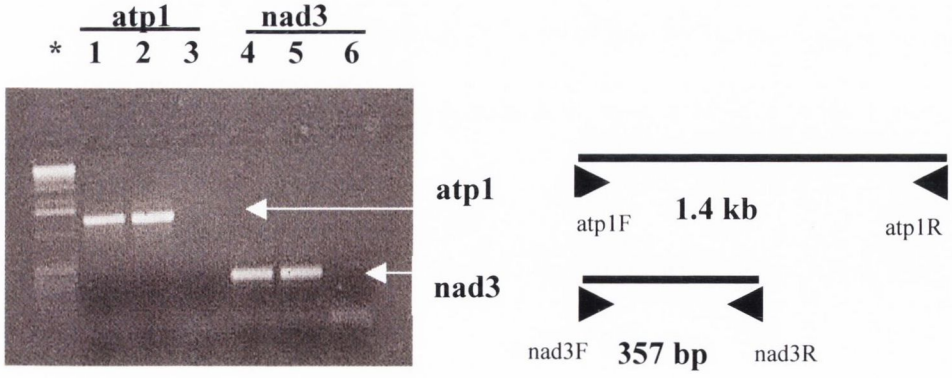
A.


cox III
rrn18S
1 2 3 4 5 6
rrn18S
1.9 kb
rrn18SF
rrn18SR
cox III
800 bp
coxIIIF
coxIIIR

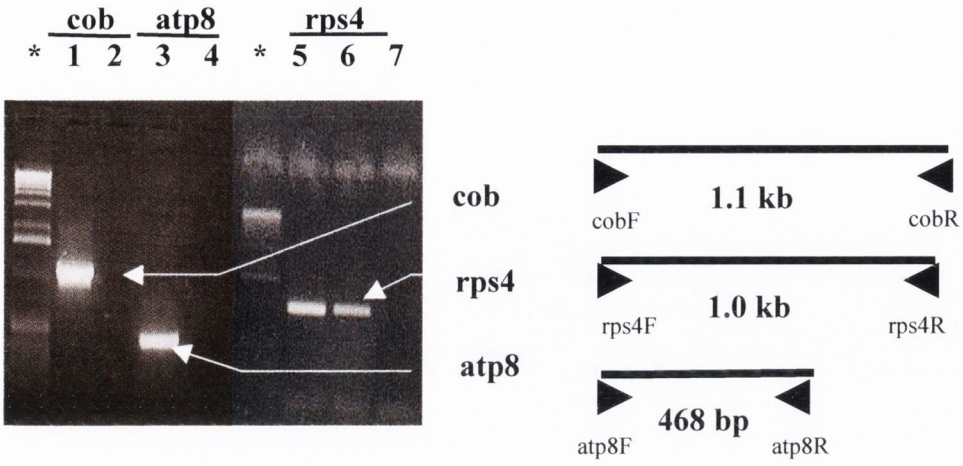
B.


rps12
cox I
* 1 2 3 4 5 6
coxI
1.5 kb
coxIF
coxIR
rps12
rps12F
378 bp
rps12R

C.


atp1
nad3
* 1 2 3 4 5 6
atp1
1.4 kb
atp1F
atp1R
nad3
nad3F
357 bp
nad3R

D.



E.

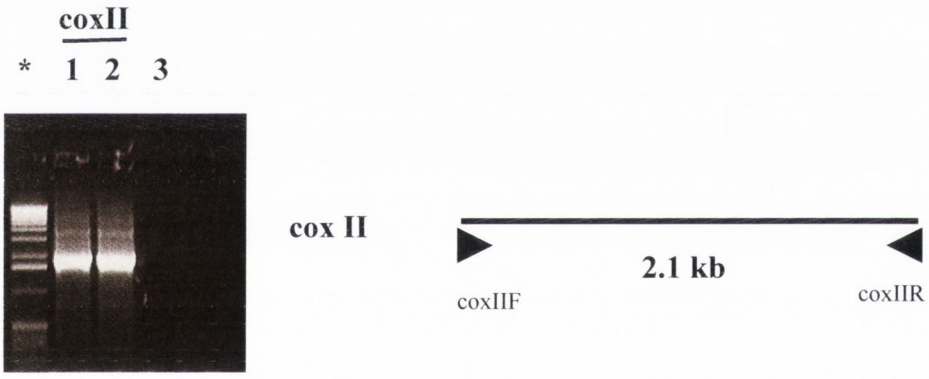


Figure 7.3 PCR products identified for 9 mitochondrial genes and a rRNA in *L.perenne.*

A. The Mitochondrial genes; Lanes 1 and 2 is the *coxIII* (800 bp) PCR fragment. Lanes 4 and 5 is the *rrn18S* (1.9 kb) PCR fragment. Lanes 3 and 6 is the negative control. **B.** The Mitochondrial genes; Lanes 1 and 2 is the *rps12* (357 bp) PCR fragment. Lanes 4 and 5 is the *coxI* (1.5 kb) PCR fragment. Lanes 3 and 6 is the negative control. **C.** The Mitochondrial genes; Lanes 1 and 2 is the *atp1* (1.4 kb) PCR fragment. ,Lanes 4 and 5 is the *rrn18S* (1.9 kb) PCR fragment. Lanes 3 and 6 is the negative control. **D.** The Mitochondrial genes; Lane 1 is the *cob* (1.1 kb) PCR fragment. Lane 3 is the *atp8* (468 bp) PCR fragment. Lanes 5 and 6 is the *rps 4* (1.0 kb) PCR fragment. Lanes 2, 4 and 7 is the the negative control. **E.** The Mitochondrial gene in lanes 1 and 2 is the *coxII* PCR fragment (2.1 kb). Lane 3 is the negative control (minus DNA template). The position of the primers used in amplification are highlighted by the symbol▶. The 1 kb ladder is represented by the lane marked with the asterisk symbol *.

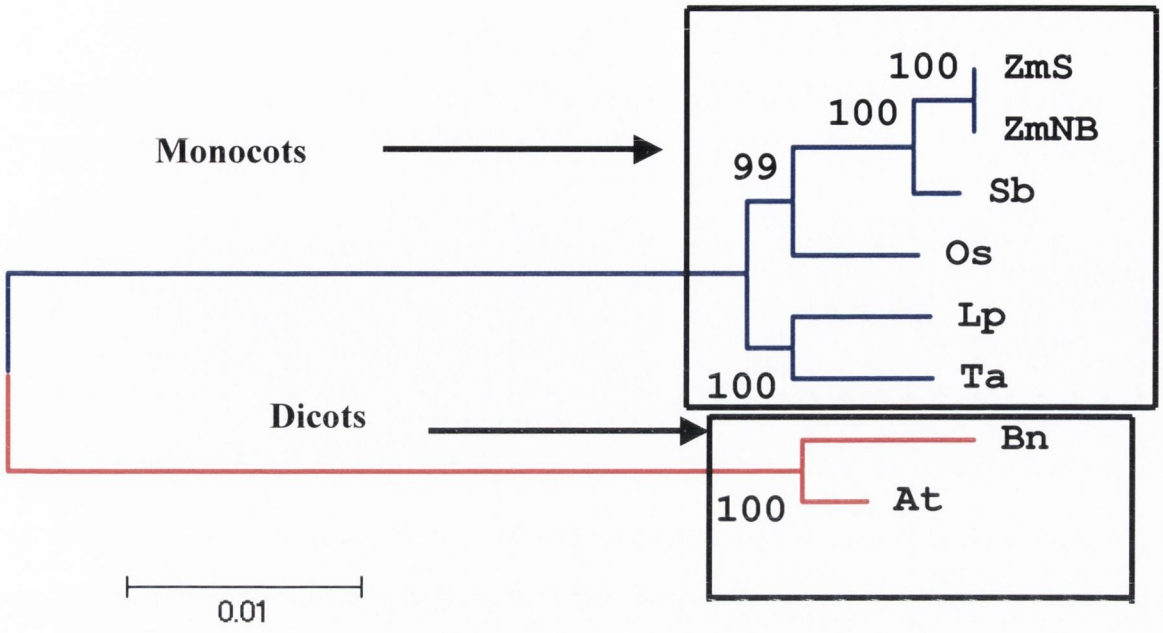


Figure 7.4 Phylogenetic relationship within the selected Monocot and Dicot species obtained from 11 Mitochondrial nucleotide coding sequences

The 11 mitochondrial sequences were joined from 3'end to 5'end yielding a 10 kb size concatenated sequence. The program ClustalW program (Thompson *et al.*, 1994) was used to align the concantenated sequence in the eight species selected. The tree drawing program MEGA ver 2.1 was used to construct a neighbor-joining tree (Saitou and Nei, 1987), using the Kimura's 2 parameter distance matrix (Kimura , 1980). The numbers above the branches are the percentage bootstrap values for 1000 replication (Felsenstein, 1985). The selected Monocot species were; *Z. mays*- ZmNB and ZmS, *S. bicolor*- Sb, *O. sativa* – Os, *L. perenne* – Lp, *T. aestivium* (Wheat)- Ta. Selected Dicot species were *A. thaliana*- At, *B. napus* – Bn.

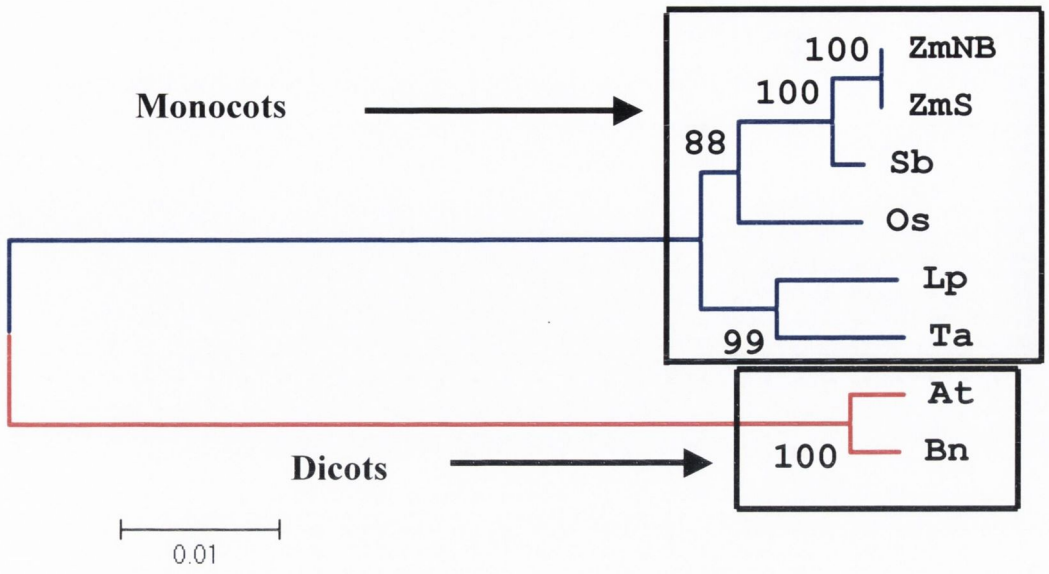


Figure 7.5 Phylogenetic relationship within the selected Monocot and Dicot species obtained from 10 Mitochondrial amino acid coding sequences

The 10 mitochondrial sequences were joined from 3' end (C-terminus) to 5' end (N-terminus) yielding a concantenated amino acid sequence. The program ClustalW program (Thompson *et al.*, 1994) was used to align the concantenated sequence in the eight species selected. The tree drawing program MEGA ver 2.1 was used to construct a neighbor-joining tree (Saitou and Nei, 1987), p distance matrix. The numbers above the branches are the percentage bootstrap values for 1000 replication (Felsenstein, 1985) The Monocot species selected; *Z. mays*- ZmNB and ZmS, *S. bicolor*- Sb, *O. sativa* – Os, *L. perenne* – Lp, *T. aestivium*- Ta. Dicot selected species *A. thaliana*- At, *B. napus* – Bn.

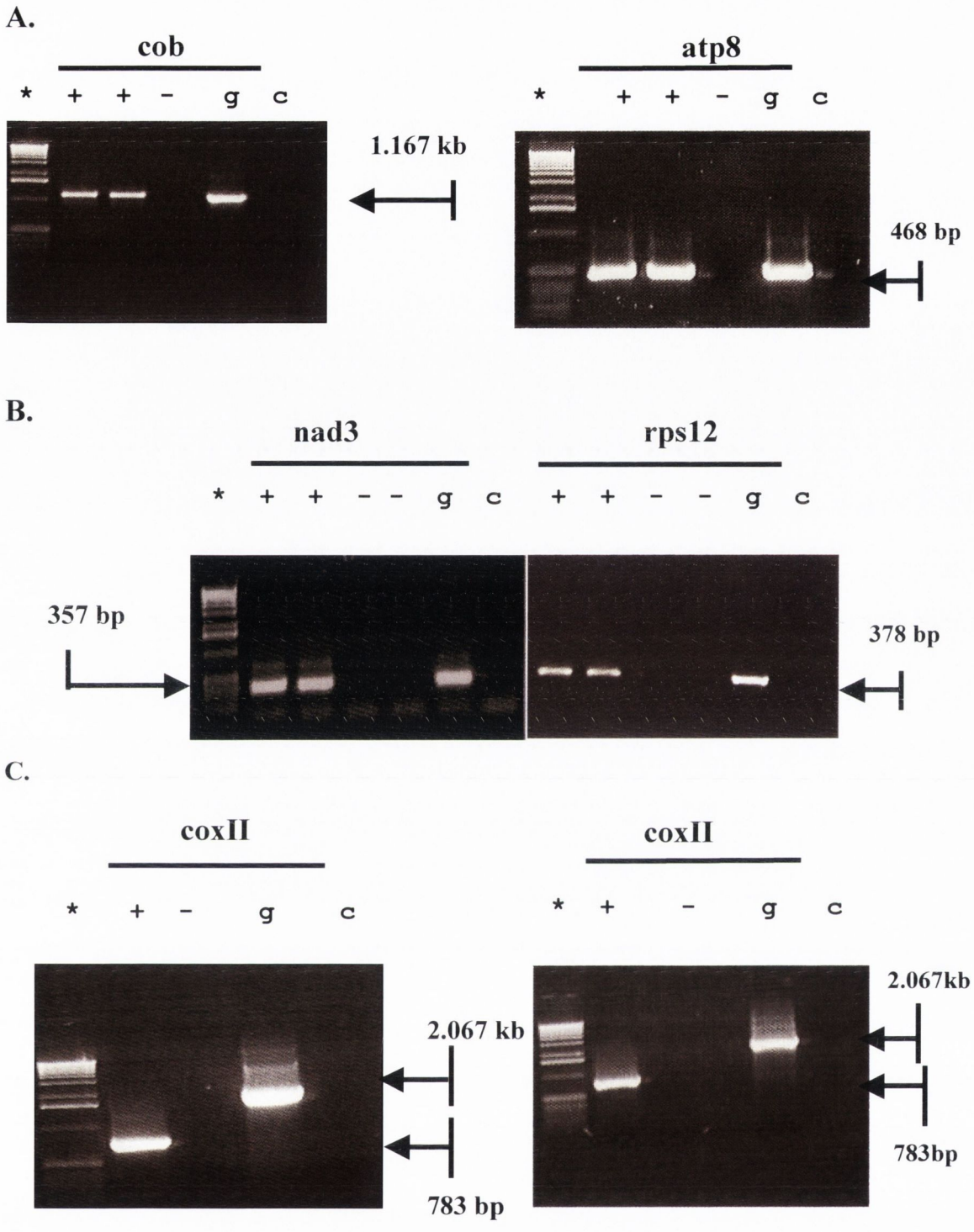


Figure 7.6 PCR amplification from single stranded cDNA from *L. perenne* (Cashel L cultivar).

A. PCR product of *cob* gene of 1.167 kb in size and *atp8* gene of 468 bp in size are both amplified from reverse transcribed cDNA (+).**B.** PCR product of *nad3* gene of 357 bp in size and *rps12* gene of 378 bp in size are both amplified from reverse transcribed cDNA (+).**C.** PCR product of *coxII* gene of 783 bp in size, amplified from reverse transcribed cDNA (+) and 2.067 kb DNA coding fragment amplified from genomic DNA template . Controls used are genomic DNA template (g), non-reverse transcribed template (-), minus DNA template (c). The 1 kb ladder is represented by the lane marked by the asterisk symbol *****.

A.

```
                          Ser     Gly      PheLeu
Lp-n          ATGATTCTTCGTTCATTAGAATGTCGATTCCTTACAATCGCTCTTTGTGATGCTGCGGAA 60
ZmNB-n        ATGATTCTTCGTTTATTAGAATGTCGATTCTTCACAATCGCTCTTTGTGATGCTGCGGAA 60
Os-n          ATGATTCTTCGTTTATTAGAATGTCGATTTCTTACAATCGCTCTTTGTGATGCTGCGGAA 60
Ta-n          ATGATTCTTCGTTCATTATCATGTCGATTTTTTACAATCGCTCTTTGTGATGCTGCGGAA 60
              ************  ****  ********  * ****************************

                                    Ala
Lp-n          CCATGGCAATTAGGATTTCAAGACGCAGCAACACCTATGATGCAAGGAATCATTGACTTA 120
ZmNB-n        CCATGGCAATTAGGATTTCAAGACGCAGCAACACCTATGATGCAAGGAATCATTGACTTA 120
Os-n          CCATGGCAATTAGGATTTCAAGACGCTGCAACACCTATGATGCAAGGAATCATTGACTTA 120
Ta-n          CCATGGCAATTAGGATTTCAAGACGCAGCAACACCTATGATGCAAGGAATCATTGACTTA 120
              ************************** ********************************

                                  Leu    Leu
Lp-n          CATCACGATATCTTTTTCTTCCTTATTCTTATTTTGGTTTTCGTATTATGGATGTTGGTT 180
ZmNB-n        CATCACGATATCTTTTTCTTCCTCATTCTGATTTTGGTTTTCGTATTATGGATGTTGGTT 180
Os-n          CATCACGATATCTTTTTCTTCCTTATTCTGATTTTGGTTTTCGTATTATGGATGTTGGTT 180
Ta-n          CATCACGATATCTTTTTCTTCCTCATTCTTATTTTGGTTTTCGTATTATGGATGTTGGTT 180
              ************************ ***** ******************************

                                                Pro
Lp-n          CGCGCTTTATGGCATTTCAACGAGCAAACTAATCCAATCCCGCAAAGGATTGTTCATGGA 240
ZmNB-n        CGCGCTTTATGGCATTTCAACGAGCAAACTAATCCAATCCCGCAAAGGATTGTTCATGGA 240
Os-n          CGCGCTTTATGGCATTTCAACGAGCAAACTAATCCAATCCCGCAAAGGATTGTTCATGGA 240
Ta-n          CGCGCTTTATGGCATTTCAACGAGCAAACTAATCCAATCCCACAAAGGATTGTTCATGGA 240
              ****************************************** ******************

                                      Ile    Pro          Leu
Lp-n          ACTACTATCGAAATTATTTGGACCATATTTCCAAGTGTCATTCTTTTGTTCATTGCTATA 300
ZmNB-n        ACTACTATCGAAATTATTTGGACCATTTTTCCTAGTGTCATTCTATTGTTCATTGCTATA 300
Os-n          ACTACTATCGAAATTATTTGGACCATATTTCCTAGTGTCATTCTTTTGTTCATTGCTATA 300
Ta-n          ACTACTATCGAAATTATTTGGACCATATTTCCAAGTGTCATTCTTTTGTTCATTGCTATA 300
              ************************** ***** *********** ****************

Lp-n          CCATCGTTTGCTCTGTTATACTCAATGGACGGGGTATTAGTAGATCCAGCCATTACTATA 360
ZmNB-n        CCATCGTTTGCTCTGTTATACTCAATGGACGGGGTATTAGTAGATCCAGCCATTACTATC 360
Os-n          CCATCGTTTGCTCTGTTATACTCAATGGACGGGGTATTAGTAGATCCAGCCATTACTATC 360
Ta-n          CCATCGTTTGCTCTGTTATACTCAATGGACGGGGTATTAGTAGATCCAGCCATTACTATC 360
              ************************************************************

                                      Thr
Lp-n          AAAGCTATTGGACATCAATGGTATTGGACTTATGAGTATTCGGACTATAACAGTTCCGAT 420
ZmNB-n        AAAGCTATTGGACATCAATGGTATTGGATTTATGAGTATTCGGACTATAACAGTTCCGAT 420
Os-n          AAAGCTATTGGACATCAATGGTATTGGACTTATGAGTATTCGGACTATAACAGTTCCGAT 420
Ta-n          AAAGCTATTGGACATCAATGGTATTGGATTTATGAGTATTCGGACTATAACAGTTCCGAT 420
              ***************************** ********************************

Lp-n          GAACAGTCACTCACTTTTGACAGTTATATGATTCCAGAAGATGATTTAGAATTGGGTCAA 480
ZmNB-n        GAACAGTCACTCACTTTTGACAGTTATATGATTCCAGAAGATGATTTAGAATTGGGTCAA 480
Os-n          GAACAGTCACTCACTTTTGACAGTTATATGATTCCAGAAGATGATTTAGAATTGGGTCAA 480
Ta-n          GAACAGTCACTCACTTTTGACAGTTATATGATTCCAGAAGATGATTTAGAATTGGGTCAA 480
              ************************************************************

Lp-n          TTACGTTTATTAGAAGTTGACAATAGAGTGGTTGTACCAGCCAAAACTCATCTACGTATG 540
ZmNB-n        TTACGTTTATTAGAAGTTGACAATAGAGTGGTTGTACCAGCCAAAACTCATCTACGTATG 540
Os-n          TTACGTTTATTAGAAGTTGACAATAGAGTGGTTGTACCAGCCAAAACTCATCTACGTATG 540
Ta-n          TTACGTTTATTAGAAGTTGACAATAGAGTGGTTGTACCAGCCAAAACTCATCTACGTATG 540
              ************************************************************

Lp-n          ATTGTAACATCCGCTGATGTACTTCATAGTTGGGCTGTACCTTCCTTAGGTGTCAAATGT 600
ZmNB-n        ATTGTAACATCCGCTGATGTACTTCATAGTTGGGCTGTACCTTCCTTAGGTGTCAAATGT 600
Os-n          ATTGTAACATCCGCTGATGTACTTCATAGTTGGGCTGTACCTTCCTTAGGTGTCAAATGT 600
Ta-n          ATTGTAACATCCGCTGATGTACTTCATAGTTGGGCTGTACCTTCCTTAGGTGTCAAATGT 600
              ************************************************************

Lp-n          GATGCTGTACCTGGTCGTTTAAATCTTACCTCCATCTTGGTACAACGAGAAGGAGTTTAC 660
ZmNB-n        GATGCTGTACCTGGTCGTTTAAATCTTACCTCCATCTTGGTACAACGAGAAGGAGTTTAC 660
Os-n          GATGCTGTACCTGGTCGTTTAAATCTTACCTCCATCTTGGTACAACGAGAAGGAGTTTAC 660
Ta-n          GATGCTGTACCTGGTCGTTTAAATCTTACCTCCATCTTGGTACAACGAGAAGGAGTTTAC 660
              ************************************************************
```

```
                                                        Pro
Lp-n            TATGGTCAGTGCAGTGAGATTTGTGGAACTAATCATGCCTTTATGCCAATCGTCGTAGAA 720
ZmNB-n          TATGGTCAGTGCAGTGAGATTTGTGGAACTAATCATGCCTTTATGCCTATCGTCGTAGAA 720
Os-n            TATGGTCAGTGCAGTGAGATTTGTGGAACTAATCATGCCTTTATGCCTATCGTCGTAGAA 720
Ta-n            TATGGTCAGTGCAGTGAGATTTGTGGAACTAATCATGCCTTTATGCCTATCGTCGTAGAA 720
                ********************************************** ************

                Ala
Lp-n            GCAGTGACTTTGAAAGATTATGCGGATTGGGTATCCAATCAATTAATCCTCCAAACCAAC 780
ZmNB-n          GCAGTGACTTTGAAAGATTATGCGGATTGGGTATCCAATCAATTAATCCTCCAAACCAAC 780
Os-n            GCGGTGACTTTGAAAGATTATGCGGATTGGGTATCCAATCAATTAATCCTCCAAACCAAC 780
Ta-n            GCAGTGACTTTGAAAGATTATGCGGATTGGGTATCCAATCAATTAATCCTCCAAACCAAC 780
                ** ************************************************************

                stop
Lp-n            TAA 783
ZmNB-n          TAA 783
Os-n            TAA 783
Ta-n            TAA 783
                ***
```

B.

```
Lp-a            MILRSLECRFLTIALCDAAEPWQLGFQDAATPMMQGIIDLHHDIFFFLILILVFVLWMLV 60
ZmNB-a          MILRLLECRFFTIALCDAAEPWQLGFQDAATPMMQGIIDLHHDIFFFLILILVFVLWMLV 60
Os-a            MILRLLECRFLTIALCDAAEPWQLGFQDAATPMMQGIIDLHHDIFFFLILILVFVLWMLV 60
Ta-a            MILRSLSCRFFTIALCDAAEPWQLGFQDAATPMMQGIIDLHHDIFFFLILILVFVLWMLV 60
                **** *.***:************************************************

Lp-a            RALWHFNEQTNPIPQRIVHGTTIEIIWTIFPSVILLFIAIPSFALLYSMDGVLVDPAITI 120
ZmNB-a          RALWHFNEQTNPIPQRIVHGTTIEIIWTIFPSVILLFIAIPSFALLYSMDGVLVDPAITI 120
Os-a            RALWHFNEQTNPIPQRIVHGTTIEIIWTIFPSVILLFIAIPSFALLYSMDGVLVDPAITI 120
Ta-a            RALWHFNEQTNPIPQRIVHGTTIEIIWTIFPSVILLFIAIPSFALLYSMDGVLVDPAITI 120
                ************************************************************

Lp-a            KAIGHQWYWTYEYSDYNSSDEQSLTFDSYMIPEDDLELGQLRLLEVDNRVVVPAKTHLRM 180
ZmNB-a          KAIGHQWYWIYEYSDYNSSDEQSLTFDSYMIPEDDLELGQLRLLEVDNRVVVPAKTHLRM 180
Os-a            KAIGHQWYWTYEYSDYNSSDEQSLTFDSYMIPEDDLELGQLRLLEVDNRVVVPAKTHLRM 180
Ta-a            KAIGHQWYWIYEYSDYNSSDEQSLTFDSYMIPEDDLELGQLRLLEVDNRVVVPAKTHLRM 180
                ********* ********************************************************

Lp-a            IVTSADVLHSWAVPSLGVKCDAVPGRLNLTSILVQREGVYYGQCSEICGTNHAFMPIVVE 240
ZmNB-a          IVTSADVLHSWAVPSLGVKCDAVPGRLNLTSILVQREGVYYGQCSEICGTNHAFMPIVVE 240
Os-a            IVTSADVLHSWAVPSLGVKCDAVPGRLNLTSILVQREGVYYGQCSEICGTNHAFMPIVVE 240
Ta-a            IVTSADVLHSWAVPSLGVKCDAVPGRLNLTSILVQREGVYYGQCSEICGTNHAFMPIVVE 240
                ************************************************************

Lp-a            AVTLKDYADWVSNQLILQTN 260
ZmNB-a          AVTLKDYADWVSNQLILQTN 260
Os-a            AVTLKDYADWVSNQLILQTN 260
Ta-a            AVTLKDYADWVSNQLILQTN 260
                ********************
```

Figure 7.7 Nucleotide (n) and amino acid (a) sequence alignment of the *coxII*-exon cDNA coding sequence in the monocot species *L. perenne* (Lp), *Z. mays* (ZmNB), *T. aestivum* (Ta), *O. sativa* (Os). A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bp is noted with the absence of the asterisk symbol *, and the corresponding codon sequence in *L. perenne* is underlined, with the abbreviated codons name indicated directly above. The TAA residues is utilised as the stop codon. **B.** Conservation of the translated *coxII* exon cDNA across the four monocot species is good with only alteration in four amino acid relative to *L. perenne* **(**highlighted in **red)** residues. The two silently edited amino acid Leu (L)-11 and Leu (L)-48 in *L. perenne* is underlined.

A.

```
                                                                    IleLeu
Lp-n        ATGCCTCAACTTGATAAATTGACTTATTTCTCACAATTCTTCTGGTTATGCATTCTCTTC 60
Os-n        ATGCCTCAACTTGATAAATTGACTTATTTCTCACAATTCTTCTGGTTATGCCTTTTCTTC 60
            ************************************************** ** *****

               Thr                 Phe
Lp-n        TTTACTTTTTATATTCTCTTATTTAATAATAATAATGGAATACTTGGAATTAGCAGAATT 120
Os-n        TTTAGTTTTTATATTCTCTTATTAAATAATAATAATGGAATACTTGGAATTAGCAGAATT 120
            **** ****************** ************************************

                                                          GlnSer
Lp-n        CTTAAACTACGGAACCAACTGCTTTCGCACCGGGGGAACAAGATCCAGAGCAAGGACCCT 180
Os-n        CTTAAACTACGGAACCAACTGCTTTCGCACCGGGGGAACAAGATCCGTTTCAAGGACCCT 180
            ********************************************    **********

                          Leu                 Thr Gly                 Ser
Lp-n        AAGAATTTGGAAGATATCTTAAGAAAAGGTTTTAGCACAGGTCTCTCATATATGTACTCT 240
Os-n        AAGAATTTGGAAGATATCTTGAGAAAAGGTTTTAGCACCGGTCTCTCATATATGTACTCC 240
            ******************** ***************** *  ****************

                                                                   Asn
Lp-n        AGTTTATCCGAAGTATCCCAATGGTGTAAGACCGTCGACTATTTGGGAAAAAGGAATAAA 300
Os-n        AGTTTATCCGAAGTATCCCAATGGTGTAAGACCGTCGACTATTTGGGAAAAAGGAGGAAA 300
            ************************************************************  ***

            Ile          Ser
Lp-n        ATAACTCTGATCTCTGATTTCGGAGAAATAAGTGGCTCACGAGGAATGGAGAGACAGATT 360
Os-n        ATCACTCTGATCTCAGATTTCGGAGAAATAAGTGGCTCACGAGGAATGGAGAGACAGATT 360
            ** *********** ********************************************

Lp-n        CTCTATTTGATCTCGAAGTCCTCATATAACACTTCTTCCAGTCGGATCACTTGTTGGAAA 420
Os-n        CTCTATTTGATCTCGAAGTCCTCATATAACACTTCTTCCAGTCGGATCACTTGTTGGAAA 420
            ************************************************************

            Asn                                                 stop
Lp-n        AACATAATGCTCACACATGTTCTACACGGGCAAGGAAGCATAATCTAA 468
Os-n        AAAATAATGCTCACACATGTTCTACACGGGCAAGGAAGCATAATCTAA 468
            ** ********************************************
```

B.

```
Lp-a        MPQLDKLTYFSQFFWLCILFFTFYILLFNNNNGILGISRILKLRNQLLSHRGNKIQSKDP 60
Os-a        MPQLDKLTYFSQFFWLCLFFFSFYILLLNNNNGILGISRILKLRNQLLSHRGNKIRFKDP 60
            ****************:::**:*****:**************************: ***

Lp-a        KNLEDILRKGFSTGLSYMYSSLSEVSQWCKTVDYLGKRNKITLISDFGEISGSRGMERQI 120
Os-a        KNLEDILRKGFSTGLSYMYSSLSEVSQWCKTVDYLGKRRKITLISDFGEISGSRGMERQI 120
            ************************************.**********************

Lp-a        LYLISKSSYNTSSSRITCWKNIMLTHVLHGQGSII 155
Os-a        LYLISKSSYNTSSSRITCWKKIMLTHVLHGQGSII 155
            ********************:**************
```

Figure 7.8 Nucleotide and amino acid sequence alignment of the *atp8* cDNA coding sequence in the monocot species *L. perenne* (Lp) and *O. sativa* (Os). A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bps is noted with the absence of the asterisk symbol *, and the corresponding codon sequence in *L. perenne* is underlined, with the abbreviated codons name indicated directly above. The TAA residues is utilized as the stop codon. **B.** Good conservation of the translated *atp8* cDNA for either monocot with an alteration of 8 amino acid residues relative to *L. perenne* (highlighted in red). The single silently edited amino acid Leu (L)-41in *L. perenne* is underlined.

A.

```
Lp-n            ATGACTATAAGGAACCAACGATTCTCTCTTCTTAAACAACCTATATACTCCACACTTAAC 60
Os-n            ATGACTATAAGGAACCAACGATTCTCTCTTCTTAAACAACCTATATACTCCACACTTAAC 60
                ************************************************************

                      Leu           Thr
Lp-n            CAGCATTTAATAGATTATCCAACCCCGAGCAATCTTAGTTATTGGTGGGGGTTCGGTTCG 120
Os-n            CAGCATTTGATAGATTATCCAACTCCGAGCAATCTTAGTTATTGGTGGGGGTTCGGTTCG 120
                ******** ************** ************************************

Lp-n            TTAGCAGGTATTTGTTTAGTCATTCAGATAGTGACTGGCGTTTTTTTAGCTATGCATTAC 180
Os-n            TTAGCAGGTATTTGTTTAGTCATTCAGATAGTGACTGGCGTTTTTTTAGCTATGCATTAC 180
                ************************************************************

Lp-n            ACACCTCATGTGGATCTAGCTTTCAACAGCGTAGAACACATTATGAGAGATGTTGAAGGG 240
Os-n            ACACCTCATGTGGATCTAGCTTTCAACAGCGTAGAACACATTATGAGAGATGTTGAAGGG 240
                ************************************************************

Lp-n            GGCTGGTTGCTCCGTTATATGCATGCTAATGGGGCAAGTATGTTTTTCATTGTGGTTTAC 300
Os-n            GGCTGGTTGCTCCGTTATATGCATGCTAATGGGGCAAGTATGTTTTTCATTGTGGTTTAC 300
                ************************************************************

Lp-n            CTTCATATTTTTCGTGGTCTATATTATGCGAGTTATAGCAGTCCTAGGGAATTTGTTTGG 360
Os-n            CTTCATATTTTTCGTGGTCTATATTATGCGAGTTATAGCAGTCCTAGGGAATTTGTTTGG 360
                ************************************************************

Lp-n            TGTCTCGGAGTTGTCATATTCCTATTAATGATTGTGACAGCTTTTATAGGATACGTACTA 420
Os-n            TGTCTCGGAGTTGTCATATTCCTATTAATGATTGTGACAGCTTTTATAGGATACGTACTA 420
                ************************************************************

Lp-n            CCTTGGGGTCAGATGAGCTTTTGGGGAGCAACAGTAATTACAAGCTTAGCTAGCGCCATA 480
Os-n            CCTTGGGGTCAGATGAGCTTTTGGGGAGCAACAGTAATTACAAGCTTAGCTAGCGCCATA 480
                ************************************************************

Lp-n            CCAGTAGTAGGAGATACCATAGTGACTTGGCTTTGGGGTGGTTTCTCCGTGGACAATGCC 540
Os-n            CCAGTAGTAGGAGATACCATAGTGACTTGGCTTTGGGGTGGTTTCTCCGTGGACAATGCC 540
                ************************************************************

Lp-n            ACCTTAAATCGTTTTTTTAGTCTTCATTATTTACTCCCCTTTATTTTAGTAGGCGCCAGT 600
Os-n            ACCTTAAATCGTTTTTTTAGTCTTCATTATTTACTCCCCTTTATTTTAGTAGGCGCCAGT 600
                ************************************************************

                     Ala
Lp-n            CTTCTTCATCTGGCCGCATTGCATCAATATGGATCAAATAATCCATTGGGTGTACATTCT 660
Os-n            CTTCTTCATCTGGCTGCATTGCATCAATATGGATCAAATAATCCATTGGGTGTACATTCT 660
                ************** ************************************************

                     Ile
Lp-n            GAGATGGATAAAATTGCTTTTTACCCTTATTTTTATGTAAAGGATCTTGTAGGTTGGGTA 720
Os-n            GAGATGGATAAAATAGCTTTTTACCCTTATTTTTATGTAAAGGATCTTGTAGGTTGGGTA 720
                ************** ************************************************

Lp-n            GCTTTTGCTATCTTTTTTTCCATTTGGATTTTTTTTGCTCCTAATGTTTTGGGGCATCCC 780
Os-n            GCTTTTGCTATCTTTTTTTCCATTTGGATTTTTTTTGCTCCTAATGTTTTGGGGCATCCC 780
                ************************************************************

Lp-n            GACAATTATATACCTGCTAATCCGATGTCCACCCCGCCTCATATTGTGCCGGAATGGTAT 840
Os-n            GACAATTATATACCTGCTAATCCGATGTCCACCCCGCCTCATATTGTGCCGGAATGGTAT 840
                ************************************************************

Lp-n            TTCCTACCGATCTATGCCATTCTTCGCAGTATACCTGACAAAGCGGGAGGTGTAGCCGCA 900
Os-n            TTCCTACCGATCTATGCCATTCTTCGCAGTATACCTGACAAAGCGGGAGGTGTAGCCGCA 900
                ************************************************************

Lp-n            ATAGCACTAGTTTTTATATCTCTCTTGGCTTTACCTTTTTTTAAAGAAATGTATGTGCGT 960
Os-n            ATAGCACTAGTTTTTATATCTCTCTTGGCTTTACCTTTTTTTAAAGAAATGTATGTGCGT 960
                ************************************************************

Lp-n            AGTTCAAGTTTTCGACCGATTTACCAAGGAATATTTTGGTTGCTTTTGGCGGATTGCTTA 1020
Os-n            AGTTCAAGTTTTCGACCGATTTACCAAGGAATATTTTGGTTGCTTTTGGCGGATTGCTTA 1020
                ************************************************************
```

```
Lp-n            CTACTAGGTTGGATCGGATGTCAACCTGTGGAGGCACCATTTGTTACTATTGGACAAATT 1080
Os-n            CTACTAGGTTGGATCGGATGTCAACCTGTGGAGGCACCATTTGTTACTATTGGACAAATT 1080
                ************************************************************
                     Val                              Pro
Lp-n            TCTTCTGTCTTTTTCTTTTTGTTCTTTGCCATAACGCCCATTCTGGGACGAGTTGGAAGA 1140
Os-n            TCTTCTTTCTTTTTCTTTTTGTTCTTTGCCATAACGCCTATTCTGGGACGAGTTGGAAGA 1140
                ****** ****************************** ********************
                                 Glu stop
Lp-n            GGAATTCCAAAATATTACACGGAATAG--------------------------- 1167
Os-n            GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCAGTCTCTTAG 1194
                ************************  **
```

B.

```
Lp-a            MTIRNQRFSLLKQPIYSTLNQHLIDYPTPSNLSYWWGFGSLAGICLVIQIVTGVFLAMHY 60
Os-a            MTIRNQRFSLLKQPIYSTLNQHLIDYPTPSNLSYWWGFGSLAGICLVIQIVTGVFLAMHY 60
                ************************************************************

Lp-a            TPHVDLAFNSVEHIMRDVEGGWLLRYMHANGASMFFIVVYLHIFRGLYYASYSSPREFVW 120
Os-a            TPHVDLAFNSVEHIMRDVEGGWLLRYMHANGASMFFIVVYLHIFRGLYYASYSSPREFVW 120
                ************************************************************

Lp-a            CLGVVIFLLMIVTAFIGYVLPWGQMSFWGATVITSLASAIPVVGDTIVTWLWGGFSVDNA 180
Os-a            CLGVVIFLLMIVTAFIGYVLPWGQMSFWGATVITSLASAIPVVGDTIVTWLWGGFSVDNA 180
                ************************************************************

Lp-a            TLNRFFSLHYLLPFILVGASLLHLAALHQYGSNNPLGVHSEMDKIAFYPYFYVKDLVGWV 240
Os-a            TLNRFFSLHYLLPFILVGASLLHLAALHQYGSNNPLGVHSEMDKIAFYPYFYVKDLVGWV 240
                ************************************************************

Lp-a            AFAIFFSIWIFFAPNVLGHPDNYIPANPMSTPPHIVPEWYFLPIYAILRSIPDKAGGVAA 300
Os-a            AFAIFFSIWIFFAPNVLGHPDNYIPANPMSTPPHIVPEWYFLPIYAILRSIPDKAGGVAA 300
                ************************************************************

Lp-a            IALVFISLLALPFFKEMYVRSSSFRPIYQGIFWLLLADCLLLGWIGCQPVEAPFVTIGQI 360
Os-a            IALVFISLLALPFFKEMYVRSSSFRPIYQGIFWLLLADCLLLGWIGCQPVEAPFVTIGQI 360
                ************************************************************

Lp-a            SSVFFFLFFAITPILGRVGRGIPKYYTE--------- 388
Os-a            SSFFFFLFFAITPILGRVGRGIPKYYTDETHRTGSVS 397
                **.*********************:
```

Figure 7.9 Nucleotide (n) and amino acid (a) sequence alignment of the *cob* cDNA coding sequence in the monocot species *L. perenne* (Lp), *O. sativa* (Os).

A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bp is noted with the absence of the asterisk symbol *, and the corresponding codon sequence in *L. perenne* is underlined, with its abbreviated name being indicated directly above. The TAG residue is utilized as the stop codon. **B.** High conservation of the translated *cob* cDNA for either monocot with an alteration of 2 amino acid residues relative to *L. perenne* (highlighted in red). The two silently edited amino acids in *L. perenne* are underlined Leu (L) –188, Phe (P) – 353 .

A.

```
                                                                  Arg
Lp-n            ATGCCTACAAAAAATCAATTGATTCGTCATGGTAGAGAAGAAAAACGGCGCACGGACCGT 60
ZmNB-n          ATGCCTACAAAAAATCAATTGATTCGTCATGGTAGAGAAGAAAAACGGCGCACGGACCGT 60
Os-n            ATGCCTACAAAAAATCAATTGATTCGTCATGGTAGAGAAGAAAAACAGCGCACGGACCGT 60
Sb-n            ATGCCTACAAAAAATCAATTGATTCGTCATGGTAGAGAAGAAAAACGGCGCACGGACCGT 60
Ta-n            ATGCCTACAAAAAATCAATTGATTCGTCATGGTAGAGAAGAAAAACGGCGCACGGACCGT 60
                ********************************************* *************

                      Ser             Gln
Lp-n            ACTCGAGCTTCGGATCAATGTCCCCAGAAGCAAGGAGTATGCCTGCGTGTTTCGACGAGA 120
ZmNB-n          ACTCGAGCTTTGGATCAATGTCCCCAGAAGCAAGGAGTATGCCTGCGTGTTTCGACGAGA 120
Os-n            ACTCGAGCTTCGGATCAATGTCCCCAAAAGCAAGGAGTATGCCTGCGTGTTTCGACGAGA 120
Sb-n            ACTCGAGCTTTGGATCAATGTCCCCAGAAGCAAGGAGTATGCCTGCGTGTTTCGACGAGA 120
Ta-n            ACTCGAGCTTTGGATCAATGTCCCCAGAAGCAAGGAGTATGCCTGCGTGTTTCGACGAGA 120
                ********** *************** ***********************************

Lp-n            ACACCGAAAAAACCTAATTCAGCTCTACGTAAGATAGCAAAAGTACGGTTGAGCAATCGA 180
ZmNB-n          ACACCGAAAAAACCTAATTCAGCTCTACGTAAGATAGCAAAAGTACGGTTGAGCAATCGA 180
Os-n            ACACCGAAAAAACCTAATTCAGCTCTACGTAAGATAGCAAAAGTACGGTTGAGCAATCGA 180
Sb-n            ACACCGAAAAAACCTAATTCAGCTCTACGTAAGATAGCAAAAGTACGGTTGAGCAATCGA 180
Ta-n            ACACCGAAAAAACCTAATTCAGCTCTACGTAAGATAGCAAAAGTACGGTTGAGCAATCGA 180
                ************************************************************

                          His                           Ser
Lp-n            CATGATATATTTGCTCACATTCCAGGCGAAGGTCATAATTCGCAGGAACATTCTATAGTC 240
ZmNB-n          CATGATATATTTGCTTACATTCCAGGCGAAGGTCATAATTTGCAGGAACATTCTATAGTC 240
Os-n            CATGATATATTTGCTCACATTCCAGGCGAAGGTCATAATTCGCAGGAACATTCTATAGTC 240
Sb-n            CATGATATATTTGCTTACATTCCAGGCGAAGGTCATAATTTGCAGGAACATTCTATAGTC 240
Ta-n            CATGATATATTTGCTTACATTCCAGGCGAAGGTCATAATTTGCAGGAACATTCTATAGTC 240
                *************** ************************ ********************

                   Val   Gly                 Ser             Ser   Arg
Lp-n            TTAGTAAGAGGAGGTAGAGTGAAAGATTCGCCAGGTGTGAAATCCCATCGTATTCGAGGA 300
ZmNB-n          TTAGTCAGAGGGGGTAGAGTGAAAGATTTGCCAGGTGTGAAATTCCATTGTATTCGAGGA 300
Os-n            TTAGTCAGAGGAGGTAGAGTGAAAGATTCGCCAGGTGTGAAATCCCATCGTATTCGAGGA 300
Sb-n            TTAGTCAGAGGAGGTAGAGTGAAAGATTTGCCAGGTGTGAAATTCCATTGTATTCGAGGA 300
Ta-n            TTAGTCAGAGGAGGTAGAGTGAAAGATTTGCCAGGTGTGAAATTCCATTGTATTCGAGGA 300
                ***** ***** **************** ************** **** ************

                Val       Leu                                           Gly
Lp-n            GTCAAGGATTTGCTGGGAATTCCGGATCGTAGAAAGGGAAGATCTAAATATGGTGCAGAA 360
ZmNB-n          GTAAAGGATTTGCTGGGAATTCCGGATCGTAGAAAGGGAAGATCTAAATATGGCGCAGAA 360
Os-n            GTCAAGGATTTACTGGGAATTCCGGATCGTAGAAAGGGAAGATCTAAATATGGTGCAGAA 360
Sb-n            GTCAAGGATTTGCTGGGAATTCCGGATCGTAGAAAGGGAAGATCTAAATATGGTGCAGAA 360
Ta-n            GTCAAGGATTTGCTGGGAATTCCGGATCGTAGAAAGGGAAGATCTAAATATGGTGCAGAA 360
                ** ******** ****************************************** ******

                   Pro   Ser   stop
Lp-n            AGACCTAAATCGAAATGA 378
ZmNB-n          AGACCAAAATCTAAATGA 378
Os-n            AGACCAAAATCGAAATGA 378
Sb-n            AGACCAAAATCGAAATGA 378
Ta-n            AGACCTAAATCGAAATGA 378
                ***** ***** ******
```

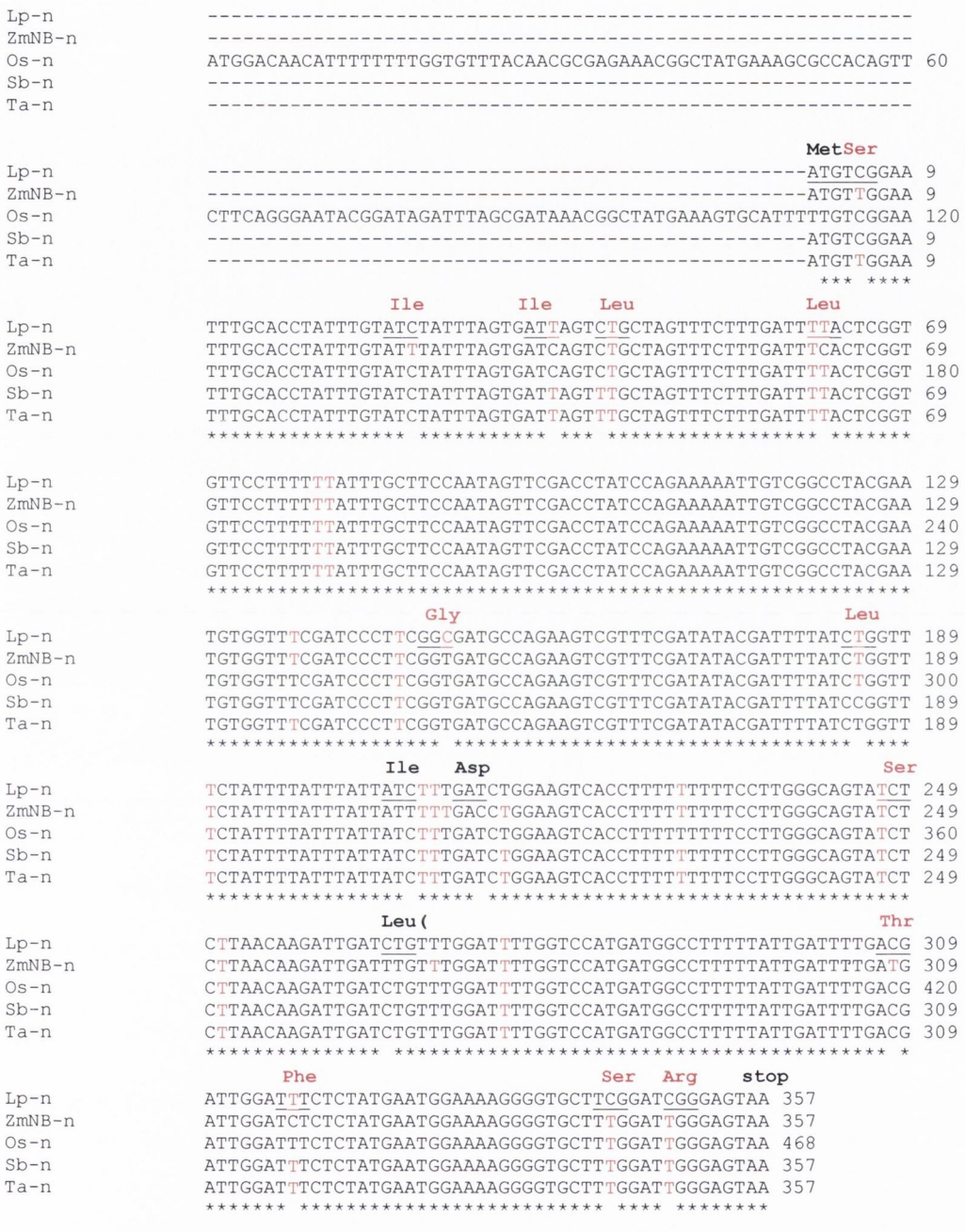
B.

```
Lp-a            MPTKNQLIRHGREEKRRTDRTRASDQCPQKQGVCLRVSTRTPKKPNSALRKIAKVRLSNR 60
ZmNB-a          MPTKNQLIRHGREEKRRTDRTRALDQCPQKQGVCLRVSTRTPKKPNSALRKIAKVRLSNR 60
Os-a            MPTKNQLIRHGREEKQRTDRTRASDQCPQKQGVCLRVSTRTPKKPNSALRKIAKVRLSNR 60
Sb-a            MPTKNQLIRHGREEKRRTDRTRALDQCPQKQGVCLRVSTRTPKKPNSALRKIAKVRLSNR 60
Ta-a            MPTKNQLIRHGREEKRRTDRTRALDQCPQKQGVCLRVSTRTPKKPNSALRKIAKVRLSNR 60
                ***************:******* ************************************

Lp-a            HDIFAHIPGEGHNSQEHSIVLVRGGRVKDSPGVKSHRIRGVKDLLGIPDRRKGRSKYGAE 120
ZmNB-a          HDIFAYIPGEGHNLQEHSIVLVRGGRVKDLPGVKFHCIRGVKDLLGIPDRRKGRSKYGAE 120
Os-a            HDIFAHIPGEGHNSQEHSIVLVRGGRVKDSPGVKSHRIRGVKDLLGIPDRRKGRSKYGAE 120
Sb-a            HDIFAYIPGEGHNLQEHSIVLVRGGRVKDLPGVKFHCIRGVKDLLGIPDRRKGRSKYGAE 120
Ta-a            HDIFAYIPGEGHNLQEHSIVLVRGGRVKDLPGVKFHCIRGVKDLLGIPDRRKGRSKYGAE 120
                *****:******* *************** **** * ************************

Lp-a            RPKSK 125
ZmNB-a          RPKSK 125
Os-a            RPKSK 125
Sb-a            RPKSK 125
Ta-a            RPKSK 125
                *****
```

Figure 7.10 Nucleotide (n) and amino acid (a) sequence alignment of the *rps12* cDNA coding sequence in the monocot species *L. perenne* (Lp), *Z. mays* (ZmNB), *O. sativa* (Os), *T. aestivum* (Ta), *S. bicolor* (Sb).

A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bp is noted with the absence of the asterisk symbol *, and the corresponding codon sequence in *L. perenne* is underlined, with the amino acid residue being indicated by its abbreviated name directly above. Alteration in amino acids due to editing are highlighted in red. The TGA residues is utilized as the stop codon. **B.** Good conservation of the translated *rps12* cDNA for either monocot with an alteration of 7 amino acid residues relative to *L. perenne* (highlighted in red).

A.


Lp-n
ZmNB-n
Os-n
Sb-n
Ta-n
Met Ser
Ile Ile Leu Leu
Gly Leu
Ile Asp Ser
Leu(Thr
Phe Ser Arg stop

B.


Lp-a
ZmNB-a
Os-a
Sb-a
Ta-a

```
Lp-a            VPFLFASNSSTYPEKLSAYECGFDPFGDARSRFDIRFYLVSILFIIFDLEVTFFFPWAVS 83
ZmNB-a          VPFLFASNSSTYPEKLSAYECGFDPFGDARSRFDIRFYLVSILFIIFDLEVTFFFPWAVS 83
Os-a            VPFLFASNSSTYPEKLSAYECGFDPFGDARSRFDIRFYLVSILFIIFDLEVTFFFPWAVS 120
Sb-a            VPFLFASNSSTYPEKLSAYECGFDPFGDARSRFDIRFYPVSILFIIFDLEVTFFFPWAVS 83
Ta-a            VPFLFASNSSTYPEKLSAYECGFDPFGDARSRFDIRFYLVSILFIIFDLEVTFFFPWAVS 83
                ************************************ ********************

Lp-a            LNKIDLFGFWSMMAFLLILTIGFLYEWKRGASDRE 118
ZmNB-a          LNKIDLFGFWSMMAFLLILMIGSLYEWKRGALDWE 118
Os-a            LNKIDLFGFWSMMAFLLILTIGFLYEWKRGALDWE 155
Sb-a            LNKIDLFGFWSMMAFLLILTIGFLYEWKRGALDWE 118
Ta-a            LNKIDLFGFWSMMAFLLILTIGFLYEWKRGALDWE 118
                ****************** ** ******** * *
```

Figure 7.11 Nucleotide (n) and amino acid (a) sequence alignment of the *nad3* cDNA coding sequence in the monocot species. *L. perenne* (Lp), *Z. mays* (ZmNB), *O. sativa* (Os), *S. bicolor (*Sb*)*, *T. aestivum* (Ta).

A. Nucleotide sequences resulting from editing (C to T transition) are highlighted in red. Any alteration in bp is noted with the absence of the asterisk symbol *, and the corresponding codon sequence in *L. perenne* is underlined, with the abbreviated codons name indicated directly above. The TAA residues is utilized for the stop codon. **B.** Good conservation of the translated *nad3* cDNA for either monocot with an alteration of 8 amino acid residues (highlighted in red) across some of the selected monocots species relative to *L. perenne* (Lp). The two silently edited amino acids Ile (I)-13, Gly (G)-50, in *L. perenne* are underlined.

Table 7.1 The percentage nucleotide identity of the 10 *L. perenne* mitochondrial protein coding sequences and a single rRNA in comparison with 5 monocot and 4 dicot plant species present within the Genbank database.

	T.aestivium	*Z.mays(NB)*	*Z.mays(S)*	*O.sativa*	*S.bicolor*	*A.thaliana*	*B.vulgaris*	*B.napus*	*N.tabacum*
coxI	99.5%	98.4%	98.4%	98.3%	98.2%	94.0%	93.2%	94.0%	94.6%
coxII	99.6%	99.1%	99.0%	99.2%	99.0%	90.9%	92.3%	90.7%	90.5%
coxIII	99.1%	99.5%	99.5%	99.5%	99.2%	95.0%	93.6%	95.6%	95.4%
atp1	99.3%	98.5%	98.1%	98.4%	98.2%	92.0%	92.6%	92.0%	92.9%
atp8	96.6%	97.7%	97.6%	96.8%	99.1%	82.6%	84.6%	82.5%	82.9%
atp9	91.6%	89.2%	88.5%	91.2%	89.3%	92.4%	88.4%	92.0%	88.4%
cob	99.8%	99.2%	99.3%	99.6%	99.3%	95.9%	94.4%	95.9%	96.3%
rps4	98.1%	97.4%	96.7%	98.2%	96.2%	89.7%	79.8%	88.9%	92.8%
rps12	99.6%	98.5%	98.5%	99.0%	99.7%	74.9%	95.2%	91.0%	94.4%
nad3	99.5%	99.5%	99.5%	99.5%	99.4%	93.0%	92.7%	93.3%	94.4%
rrn18S	99.7%	99.6%	99.3%	99.7%	99.9%	93.9%	98.4%	91.0%	95.3%
Average	98.4%	97.9%	97.7%	98.1%	97.9%	90.4%	91.4%	91.5%	92.5%

Table 7.2a Is a distance matrix showing the number of substituions per nucleotide sites for the 10 different genes and a single rRNA in eight different plant species by the Kimura -2- parameter (Kimura,1980) using the MEGA ver 2.1 program (Kumar *et al*, 2001). Above the diagonal of the matrix (in red text) are the standard error values estimated.

	Bn	At	Lp	Os	ZmS	ZmNB	Sb	Ta
Bn		[0.00102]	[0.00294]	[0.00293]	[0.00299]	[0.00299]	[0.00298]	[0.00295]
At	0.01014		[0.00287]	[0.00286]	[0.00289]	[0.00289]	[0.00289]	[0.00287]
Lp	0.07922	0.07542		[0.00125]	[0.00133]	[0.00133]	[0.00131]	[0.00111]
Os	0.07855	0.07487	0.01526		[0.00115]	[0.00115]	[0.00113]	[0.00125]
ZmS	0.08126	0.07667	0.01722	0.01300		[0.00000]	[0.00069]	[0.00135]
ZmNB	0.08126	0.07667	0.01722	0.01300	0.00000		[0.00069]	[0.00135]
Sb	0.08090	0.07642	0.01681	0.01259	0.00465	0.00465		[0.00132]
Ta	0.07956	0.07532	0.01218	0.01536	0.01784	0.01784	0.01701	

Plant Species nucleotide

Bn	*Brassica napus*
At	*Arabidopsis thaliana*
Lp	*Lolium perenne*
Os	Oryzae sativa
ZmS	*Zea mays* (S)
ZmNB	*Zea mays (NB)*
Sb	*Sorghum bicolor*
Ta	*Triticum aestivum*

Table 7.2b Is a distance matrix according to the p-distance value showing the number of differences per sites for 10 different genes in eight different plant species, using the MEGA ver2.1 program (Kumar *et al*, 2001). Above the diagonal of the matrix (in red text) are the standard error values estimated.

	ZmNB	ZmS	Sb	Os	Lp	Ta	At	Bn
ZmNB		[0.00000]	[0.00136]	[0.00247]	[0.00289]	[0.00292]	[0.00586]	[0.00586]
ZmS	0.00000		[0.00136]	[0.00247]	[0.00289]	[0.00292]	[0.00586]	[0.00586]
Sb	0.00509	0.00509		[0.00236]	[0.00285]	[0.00283]	[0.00584]	[0.00584]
Os	0.01707	0.01707	0.01562		[0.00283]	[0.00287]	[0.00583]	[0.00582]
Lp	0.02361	0.02361	0.02288	0.02252		[0.00236]	[0.00590]	[0.00586]
Ta	0.02397	0.02397	0.02252	0.02325	0.01562		[0.00590]	[0.00590]
At	0.10570	0.10570	0.10498	0.10461	0.10716	0.10716		[0.00154]
Bn	0.10570	0.10570	0.10498	0.10425	0.10570	0.10716	0.00654	

Plant Species amino acid

ZmNB	*Zea mays*(NB)
ZmS	*Zea mays*(S)
Sb	*Sorghum bicolor*
Os	*Oryza sativa*
Lp	*Lolium perenne*
Ta	*Triticum aestivum*
At	*Arabidopsis thaliana*
Bn	*Brassica napus*

Table 7.3 Number of nucleotide substitutions per nucleotide site (K) for 10 mitochondrial genes and a single rRNA in a concatenated array

Taxon pair	K	Estimated Divergence Time (Myr)
Maize - Wheat	0.01784 ± 0.00125	50.0 – 70.0
Maize – Rice	0.01300 ± 0.00115	36.4 – 51.0
Maize - Sorghum	0.00465 ± 0.00069	13.0 – 18.2
Wheat - Lolium	0.01218 ± 0.00111	34.2 – 47.8
Monocots –Dicots	0.07801 ± 0.00289	218.6 – 306.1
Brassica – Arabidopsis	0.01014 ± 0.00102	28.4 – 39.7

Maize vs Wheat ***(K/2T)***

K value is based on Maize vs Wheat (K = 0.01784)

T is estimated divergence time of 50-70 Myr

Rate of 0.127 - 0.178 x10^{-9} Nucleotide substitution per site per year.

Table 7.4 (a): Number of Synonymous (Ks) substitutions per site in mitochondrial DNA using the method by Li *et al.*, (1985) in MEGA (ver 2.1) for 9 mitochondrial genes.

GENES	*Zea mays-Triticum aestivum*	*Zea mays – Oryzae sativa*	*Triticum aestivum –Lolium perenne*	*Monocot-Dicot*
coxI	0.04787 ± 0.01207	0.02364 ± 0.00841	0.03581 ± 0.01042	0.17112 ± 0.02451
coxII	0.03050 ± 0.01373	0.02431 ± 0.01221	0.01811 ± 0.01049	0.15979 ± 0.03279
coxIII	0.00552 ± 0.00557	0.01099 ± 0.0074	0.01095 ± 0.0079	0.12054 ± 0.0026
atp1	0.05693 ± 0.01320	0.05581 ± 0.0135	0.03238 ± 0.0098	0.23718 ± 0.02908
atp8	0.06730 ± 0.02816	0.01045 ± 0.01094	0.09047 ± 0.03042	0.39201 ± 0.07990
cob	0.01999 ± 0.00898	0.02410 ± 0.00988	0.00396 ± 0.00397	0.09029 ± 0.02015
rps4	0.06184 ± 0.01745	0.03042 ± 0.01202	0.02334 ± 0.01055	0.18350 ± 0.03352
rps12	0.05996 ± 0.02725	0.07109 ± 0.02942	0.01146 ± 0.01151	0.22388 ± 0.06519
rrn18S	0.00563 ± 0.00399	0.00563 ± 0.00399	0.00281 ± 0.00281	0.04423 ± 0.01240
Weighted average	0.03950 ± 0.01448	0.02840 ± 0.01197	0.02547 ± 0.00970	0.18028 ± 0.03334

Table 7.4 (b): Time divergence (Myr) estimated from a Ratio value obtained from the average number of Synonymous (Ks) substitutions per site relative to Maize-Wheat.

Time	Maize-Wheat	Maize-Rice	Wheat-Lolium	Monocot-Dicot
Ratio of synonymous Substitution Number	---------	0.7188	0.6448	4.564
Landmark Time (Myr)	50 -70	---------	---------	---------
Estimated DivergenceTime (Myr)	---------	35.9 – 50.3	32.2 – 45.1	228.2 – 319.5

Maize vs Wheat ***(Ks/2T)***

Ks value is based on Maize vs Wheat split (Ks = 0.18028)

T is estimated divergence time of 50-70 Myr

Rate of 0.28 - 0.39x10^{-9} synonymous nucleotide substitutions per site per year

Table 7.5A Comparison of the percentage nucleotide identity in the cDNA coding sequence from *L. perenne* with cDNA in plant species of *Z. mays* (NB), *O. sativa*, *T. aestivum*, *S. bicolor*, *A. thaliana*, *B. napus*, *B.vulgaris*. The dash symbol – represents gene whose cDNA sequence was not available.

	Zea mays (NB)	*Oryzae sativa*	*Triticum aestivum*	*Sorghum bicolor*	*Arabidopsis thaliana*	*Brassica napus*	*Beta vulgaris*
coxII	98.72%	99.11%	98.59%	–	91.60%	91.82%	93.11%
atp8	–	96.79%	–	–	–	83.76%	86.12%
atp9	92.0%	93.34%	94.23%	92.88%	94.22%	93.77%	90.67%
cob	–	99.31%	–	–	97.0%	97.0%	95.63%
rps12	96.82%	98.94%	98.15%	97.88%	89.95%	90.47%	94.71%
nad3	96.64%	98.88%	98.60%	98.60%	93.55%	94.12%	94.40%
Average	**96.05%**	**97.73%**	**97.39%**	**96.45%**	**93.26%**	**91.82%**	**92.44%**

Table 7.5B Comparison of the percentage amino acid identity in the cDNA coding sequence from *L. perenne* with cDNA in plant species of *Z. mays* (NB), *O. sativa*, *T. aestivum*, *S. bicolor*, *A. thaliana*, *B. napus*, *B. vulgaris*. The dash symbol – represents genes whose cDNA amino acid sequence was not available.

	Zea mays (NB)	*Oryzae sativa*	*Triticum aestivum*	*Sorghum bicolor*	*Arabidopsis thaliana*	*Brassica napus*	*Beta vulgaris*
coxII	98.85%	99.61%	99.23%	–	90.77%	88.85%	90.77%
atp8	–	94.84%	–	–	–	76.78%	77.50%
atp9	100.0%	100%	100.0%	100%	100.%	100%	100%
cob	–	99.48%	–	–	97.43%	96.91%	95.11%
rps12	95.02%	99.20%	95.20%	94.20%	92.80%	87.20%	91.20%
nad3	95.0%	97.46%	97.50%	97.46%	97.46%	94.92%	95.76%
Average	**97.21%**	**98.20%**	**97.98%**	**97.22%**	**93.33%**	**90.77%**	**91.72%**

Table 7.6 RNA editing sites in *L. perenne* illustrating the number of silent and non-silent editing sites in several mitochondrial genes.

Gene	Coding (bp)	Number of Editing sites (C→T)	(T→C)	Non-Silent edits	Silent edits
Complex I					
nad3	357	17	1	13	2
Complex III					
cob	1,167	19	0	17	2
Complex IV					
cox II	783	18	0	15	2
Complex V					
atp8	468	4	0	3	1
*atp9**	225	0	0	0	0
Ribosomal					
rps12	378	0	0	0	0
Total	**3,379**	**58**	**1**	**48**	**7**

Table 7.7 The location and the distinction of silent (highlighted in red) and non-silent editing sites (highlighted in black) in *L. perenne* for the coding genes (i) coxII (ii) atp8 (iii) cob (iii) nad3. The asterisk symbol * denotes the uncommon T to C transition, whereas the common C to T transition is left unmarked.

(i) *coxII*

Codon Number	*L. perenne* cDNA	*L. perenne* DNA
11	CTT(L)	CTC
26	TTT(F)	TCT(S)
48	CTT(L)	CTC
56	TTA(L)	TCA(S)
57	TGG(W)	CGG(R)
87	TGG(W)	CGG(R)
129	TGG(W)	CGG(R)
150	ATG(M)	ACG(T)
156	TTA(L)	CCA(P)
161	TTA(L)	TCA(S)
184	TCC(S)	CCC(P)
188	CTT(L)	CCT(P)
196	TTA(L)	TCA(S)
207	TTA(L)	TCA(S)
303	TTG(L)	TCG(S)
318	TGT(C)	CGT(R)
325	ATG(M)	ACG(T)

(ii) *atp8*

Codon Number	*L. perenne* cDNA	*L. perenne* DNA
20	TTC(F)	CTC(L)
41	CTT(L)	CTC
67	TTA(L)	TCA(S)
148	CTA(L)	CCA(P)

(iii) *cob*

Codon Number	*L. perenne* cDNA	*L. perenne* DNA
60	TAC(Y)	CAC(H)
96	TTC(F)	CTC(L)
100	TAC(Y)	CAC(H)
109	TAT(Y)	CAT(H)
120	TGG(W)	CGG(R)
140	CTA(L)	CCA(P)
188	CTT(L)	CTC
190	TAT(Y)	CAT(H)
194	TTT(F)	CTT(L)
227	TTT(F)	TCT(S)
239	TGG(W)	CGG(R)
242	TTT(F)	TCT(S)
270	TCC(S)	CCC(P)
285	TAT(Y)	CAT(H)
303	CTA(L)	CCA(P)
328	TAC(Y)	CAC(H)
352	TTT(F)	TTC
361	TCT(S)	CCT(P)
375	CTG(L)	CCG(I)

(iv) *nad3*

Codon Number	*L. perenne* cDNA	*L. perenne* DNA
13	ATT(I)	ATC
15	CTG(L)	CCG(P)
21	TTA(L)	CCA(P)
27	TTA(L)	CCA(P)
46	TTC(F)	TCC(S)
49	TTC(F)	TCC(S)
50*	GGC(G)	GGT
62	CTG(L)	CCG(P)
64	TCT(S)	CCT(P)
70	TTT(F)	CCT(P)
77	TTT(F)	TCT(S)
83	TCT(S)	CCT(P)
84	CTT(L)	CCT(P)
92	TTT(F)	TCT(S)
106	TTT(F)	TCT(S)

Table 7.8 Illustrates the number of editing sites in *L. perenne* relative to selected monocot and dicot plant species. The Red box highlights monocots in which edited sites have been determined for particular genes (highlighted in *black type*), and the number of editing sites located at the same position in the same gene of a monocot species relative to *L. perenne* is shown in red type (enclosed in parentheses).

	Lolium perenne	*Zea mays(NB)*	*Oryzae sativa*	*Triticum aestivum*	*Sorghum bicolor*	*Arabidopsis thaliana*	*Brassica napus*	*Beta vulgaris*
coxII	18	19 (15)	19 (17)	16 (14)	–	15	13	9
atp8	4	–	4 (4)	–	–	–	3	2
atp9	0	7 (0)	8 (0)	8 (0)	8 (0)	4	4	5
cob	19	–	19 (16)	–	–	7	7	13
rps12	0	6 (0)	0	6 (0)	6 (0)	8	7	6
nad3	18	22 (14)	15 (13)	21 (16)	19 (15)	11	10	12

CHAPTER 8

General Discussion and Future Perspectives

8.1 The Dpo- and Rpo-like sequences in *L. perenne*

The 9.6 kb CMSi region uniquely associated with the mt genome of CMS *L. perenne* has features that are typical of integrated mitochondrial linear plasmids found in fungi and higher plants e.g. *C. purpurea* and *D. carota,* respectively (Robison and Wolyn, 2005). Based on sequence similarity, ORF1 and ORF2, which comprise most of the element, encode a family B Dpo and single-subunit Rpo, respectively. ORF3 encodes an unknown protein that may play a role in plasmid maintenance, as suggested by Handa (2002), and Robison and Wolyn (2005). Surprisingly, all three ORFs were also identified by PCR in two fertile lines of *L. perenne*. This would seem to suggest that CMSi-like sequences exist at very low copy number in certain fertile lines, either as free-living linear plasmids or as sub-genomic mtDNA molecules. Although phylogenetic analyses indicate a close affiliation with linear plasmid-like sequences of higher plants, the origin of CMSi remains a mystery. This might be resolved by investigating, initially by PCR, the presence of CMSi-like sequences in a large number of *L. perenne* and *F. pratensis* accessions and also a range of species from each of these genera. Proteinase K treatment of mtDNA-enriched fractions followed by Southern blot analysis could then be used to identify lines containing integrated or high copy number CMSi sequences. These studies might also be extended to include species of the fungal genera *Epichloe* and *Acremonium* (Byrd *et al.*, 1990; Schardl *et al.*, 2004), the mycelia of which are known to establish close associations with the reproductive organs of *Lolium* and *Festuca* species, and which might conceivably have lead to a rare horizontal transmission of plasmids between symbiont and host.

So far, the enzymatic properties of the polymerases encoded by plant mt linear plasmids have not yet been investigated. However, Kempken *et al.,* (1989) showed that the protein-primed Dpo encoded by the pAI2 plasmid of *Ascobolus immerus* can incorporate radioactively labeled nucleotides at the ends of the plasmid and into the appropriate strand. We attempted to express fragments of the

LpDpo protein in *E. coli*, but the proteins were either insoluble or could not be purified to homogeneity. Also, the antibodies raised against insoluble recombinant LpDpo did not detect LpDpo expression in mitochondria of the CMS line. Therefore, in terms of future work it might be worth pursuing these goals using a different strategy to produce recombinant protein and better anti-LpDpo and anti-LpRpo antibodies. For example, smaller sized fragments derived from either polymerase might be used to generate a series of antibodies, with an emphasis on protein sites encompassing predicted surface epitopes. Expression of the full-length proteins in a eukaryotic expression system (e.g. the yeast *Pichia* or an insect cell line) might also be considered which might overcome the commonly encountered problem of recombinant protein insolubility in *E. coli*.

8.2 The association of *orfC9* with CMS in *L. perenne*

Whether or not *orfC9* plays a contributory role in CMS in *L. perenne* has not yet been resolved. If this CMS-associated gene is active only in tapetal cells or young microspores, a full analyses of its role in CMS will be challenging due to limitations on the availabilty of tissue. No alterations in processing or in the amount of the *atp9/orfC9* co-transcript were observed in leaf tissue from fertility-restored lines. Similarly, there was no alteration in *atp9/orfC9* transcripts by RNA editing which suggests that editing does not play a role in fertility restoration. It is therefore possible that *orfC9* plays no role in CMS in *L. perenne,* or alternatively that restoration acts at a translational or post-translational level. In the *B. napus* Ogura and *Kosena* radish CMS systems, the *orf138-atp8* and *orf125-atp8* co-transcript, respectively, were unaltered in lines restored to fertility, but the restorer genes acted post-translationally to destabilise the ORF138 Ogura and ORF125 Kosena proteins (Bellaoui *et al.*, 1999; Koizuka *et al.*, 2000, 2003). Unfortunately, we were unable to detect ORFC9 protein in leaf tissue using polyclonal antibodies generated from recombinant ORFC9 protein produced in *E. coli*. Thus, it was not possible to investigate whether restoration to fertility was associated with reduced ORFC9 levels. Future work should ideally use immunohistochemical methods and a high titre anti-ORFC9 antibody to investigate ORFC9 levels specifically in anthers of sterile and restored lines. Such

an approach would have the added advantage of localizing ORFC9 expression to particular tissues and/or cell types.

8.3 Disintegration of the tapetum and pollen abortion in CMS *L. perenne*

According to previous studies, the pollen coat of sterile microspores lack key components e.g. sporopollenin, intine and lipidic compounds, the synthesis of which depends on the tapetum. Thus, the premature arrest of tapetal development can lead to pollen abortion by depriving microspores of critical metabolites required for their development and maturation. In CMS *L. perenne,* the tapetal layer begins to appear abnormal at stage 5 when microspores are released from tetrads and tapetal vacuolation becomes apparent. We do not know if this process is evidence of programmed cell death (PCD) or simple necrotic disintegration of tapetal cells. Nor do we know precisely how tapetal cell death leads to pollen abortion except that, by definition, mitochondrial dysfunction might play a role. Future studies should therefore focus on (a) mitochondrial integrity during anther/microspore development and (b) strategies to determine whether PCD is a feature of tapetal cell disintegration in *L. perenne*. The number and integrity of mitochondria could be monitored using confocal microscopy and the fluorescent vital stain MitoTracker (Invitrogen) which specifically accumulates in mitochondria that are metabolically intact (i.e. maintain a membrane potential). The occurrence of necrotic cell death vs. apoptotic cell death (PCD) could be investigated using the ISNT (*In Situ* Nick Translation) and TUNEL (**T**dT-mediated X-d**U**TP **n**ick **e**nd **l**abeling) assays, respectively. Each of these assays relies on the incorporation of fluorescent labels into nicked chromosomal DNA, which can be visualized by fluorescence microscopy.

8.4 *L. perenne* mitochondrial genes as an evolutionary guide

In this study the nucleotide sequence of 10 *L. perenne* mitochondrial genes was determined and RNA editing was examined for 6 of these. The editing frequency was high for the majority of the respiratory complex genes except for *atp9* which showed no evidence of editing. In general, editing increased the level of amino acid sequence conservation between the *L. perenne* predicted COXII,

ATP8, COB, NAD3 proteins and those of other higher plant species. To obtain a more complete picture of RNA editing in *L. perenne*, multiple cDNA clones for each of the mt protein-coding genes should be obtained and sequenced in order to identify both completely and partially edited sites in these genes. In time, such data should allow us to gain a deeper insight into the evolutionary relationships between *L. perenne* and related members of the flowering plant kingdom.

The mitochondrial coding sequence data generated in this study was used to estimate the time at which *L. perenne* diverged from its last common ancestor and to examine its phylogenetic position within the family *Gramineae*. Numbers of nucleotide substitution based on coding sequence differences and synonymous substitutions were separately examined and were found to generate similar divergence times. For example, the number of synonymous nucleotide substitution per site (Ks) in coding sequences seem to indicate that the genus *Lolium* may have diverged from *Triticum* 32 - 45 Myr ago. A more accurate phylogenetic analysis of *L. perenne* in relation to other members of the *Gramineae* (e.g. those whose mt genome has been fully sequenced) could be undertaken if the remaining *L. perenne* mt protein coding genes (ca. 20 - 25 genes) were sequenced.

References

Aarts, MG., Hodge, R., Kalantidis, K., Florack, D., Wilson, ZA., Mulligan, BJ., Stiekema, WJ., Scott, R., Pereira, A. (1997). The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *Plant J.* **12:** 615-623.

Abad, AR., Mehrtens, BJ., Mackenzie, SA. (1995). Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell* **7:** 271-285.

Abdelnoor, RV., Yule, R., Elo, A., Christensen, AC., Meyer-Gauen, G., Mackenzie, SA. (2003). Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc. Natl. Acad. Sci. U S A* **100:** 5968-73.

Adams, KL., Qiu, YL., Stoutemyer, M., Palmer, JD. (2002). Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. USA* **99**: 9905-9912.

Akagi, H., Shimada, H., Fujimara, T. (1995). High frequency inter-parental recombination between mitochondrial genomes of rice cybrids. *Curr. Genet.* **29:** 58-65.

Altschul, SF., Madden, TL., Schaffer, AA., Zhang, JH., Zhang, Z., Miller, W., Lipman, DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

Arabidopsis Genome Initiative. (2000). Analysis of the Genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408:** 796-815.

Arganoza, MT., Min, J., Hu, Z., Atkins, RA. (1994). Distribution of seven homology groups of mitochondrial plasmids in *Neurospora*: evidence for widespread mobility between species in nature. *Curr. Genet.* **26:** 62-73.

Ariizumi, T., Hatakeyama, K., Hinata, K., Sato, S., Kato, T., Tabata, S., Toriyama, K. (2003). A novel male sterile mutant of *Arabidopsis thaliana*, faceless pollen-1, produces

pollen with a smooth surface and an acetolysis –sensitive exine. *Plant Mol. Biol.* **53:** 107-116.

Ariizumi, T., Hatakeyama, K., Hinata, K., Inatsugi, R., Nishida, I., Sato, S., Kato, T., Tabata, S., Toriyama, K. (2004). Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. *Plant J.* **39**: 170-181.

Backert, S., Dorfel, P., Lurz, R., Borner, T. (1996a). Rolling-circle replication of mitochondrial DNA in the higher plant *Chenopodium album* (L.). *Mol. Cell Biol.* **16:** 6285-6294.

Backert, S., Lurz, R., Borner ,T. (1996b). Electron microscopic investigation of mitochondrial DNA from *Chenopodium album* (L.). *Curr Genet.* **29:** 427-436.

Backert, S., Meissner, K., Borner, T. (1997a). Unique features of the mitochondrial rolling circle-plasmid mp1 from the higher plant *Chenopodium album* (L.). *Nucleic Acid Res.* **25:** 582-589.

Backert, S., Nielsen, BL., Borner, T. (1997b). The mystery of the rings: structure and replication of mitochondrial genomes from higher plant. *Trends Plant Sci.* **2:** 477-483.

Backert, S. and Borner, T. (2000). Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). *Curr. Genet.* **37:** 304-314.

Balfourier, F., Charmet, G., Ravel, C. (1998). Genetic differentiation within and between natural populations of perennial and annual ryegrass (*Lolium perenne* and *L. rigidum*). *Heredity* **81:** 100-110.

Balk, J. and Leaver, CJ. (2001). The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. *Plant Cell* **13:** 1803-1818.

Barrena, GV. and Wilson, ZA. (2006). Altered tapetal PCD and pollen wall development in the *Arabidopsis ms1* mutant. *J. Exp. Bot.* **57:** 2709-2717.

Beals, TP. and Goldberg, RB. (1997). A novel cell ablation strategy blocks tobacco anther dehiscence. *Plant Cell* **4:** 879-887.

Bedinger, P. (1992). The Remarkable Biology of Pollen. *Plant Cell* **4**: 879-887.

Bellaoui, M., Grelon, M., Pelletier, G., Budar, F. (1999). The restorer *Rfo* gene acts post-translationally on the stability of the ORF138 Ogura CMS-associated protein in reproductive tissues of rapeseed cybrids. *Plant Mol. Biol.* **40:**893-902.

Bendich, AJ. (1993). Reaching for the circle: the study of mitochondrial genome structure. *Curr. Genet.* **24:** 279-290.

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

Benslimane, AA., Rode, A., Quétier, F., Hartmann, C. (1994). Characterization of two minicircular plasmid-like DNAs isolated from date-palm mitochondria. *Curr. Genet.* **26:** 535-541.

Bentolila, S., Alfonso, AA., Hanson, MR. (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. USA* **99:** 10887-10892.

Bergman, P., Edqvist, J., Farbos, I., Glimelius, K. (2000). Male-sterile tobacco displays abnormal mitochondrial *atp1* transcript accumulation and reduced floral ATP/ADP ratio. *Plant Mol. Biol.* **42:** 531-544.

Bernad, A., Zabollos, A., Salas, M., Blanco, L. (1987). Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. *EMBO J.* **6:** 4219-4225.

Bernad, A., Blanco, L., Lazaro, JM., Martin, G., Salas, M. (1989). A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* **59:** 219-228.

Bevan, M. (1984). Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Res.* **12:** 8711–8721.

Binder, S., Marchfelder, A., Brennicke, A. (1994). RNA editing of tRNA(Phe) and tRNA(Cys) in mitochondria of *Oenothera berteriana* is initiated in precursor molecules. *Mol. Gen. Genet.* **244:** 67-74.

Blaisonneau, J., Nosek, J., Fukuhara, H. (1999). Linear DNA plasmid pRK2 of *Pichia kluyveri*: Distinction between cytoplasmic and mitochondrial linear plasmids in yeast. *Yeast* **15:** 781-791.

Blanco, L., Bernad, A., Blasco, MA., Salas, M. (1991). A gene structure for DNA-Dependant DNA polymerases. *Gene* **100:** 27-38.

Boeshore, ML., Hanson, MR., Izhar, S. (1985). A variant mitochondrial DNA arrangement specific to Petunia sterile somatic hybrids. *Plant Mol. Biol.* **4:** 125-132.

Bois, F., Barroso, G., Gonzalez, P., Labarere, J. (1999). Molecular cloning, sequence and expression of Aa-polB, a mitochondrial gene encoding a family B DNA polymerase from the edible basidiomycete *Agrocybe aegerita. Mol Gen. Genet.* **261:** 508-513.

Bonhomme, S., Budar, F., Lancelin, D., Small, I., Defranc, MC., Pelletier, G. (1992). Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in *Brassica* cybrids. *Mol. Gen. Genet.* **235:** 340-348.

Bonnema, AB., Castillo, C., Reiter, N., Cunningham, M., Adams, HP., O'Connell, M. (1995). Molecular and ultrastructural analysis of a nonchromosomal variegated mutant. Tomato mitochondrial mutants that cause abnormal leaf development. *Plant Physiol.* **109:** 385-392.

Bonos, SA., Clarke, BB., Meyer, WA. (2006). Breeding for disease resistence in the major cool-season turfgrasses. *Annu. Rev. Phytopathol.* **44:** 213-234.

Boore, JL. (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* **27:** 1767-1780.

Boutry, M., Nagy, F., Poulsen, C., Aoyagi, K., Chua, N-H. (1987). Targeting of bacterial chloramphenicol acetyltransferase to mitochondria in transgenic plants. *Nature* **328:** 340-

342.

Brandt, P., Sunkel, S., Unseld, M., Brennicke, A., Knoop, V. (1992). The nad4L gene is encoded between exon c of nad5 and orf25 in the *Arabidopsis* mitochondrial genome. *Mol. Gen. Genet.* **236:** 33-38.

Bras, M., Queenan, B., Susin, SA. (2005). Programmed cell death via mitochondria: different modes of dying. *Biochemistry* **70:** 231-239.

Bremer, K. (2000). Early Cretaceous lineages of monocot flowering plants. *Proc. Nat. Acad. Sci. USA* **97:** 4707-4711.

Brown, GG. and Zhang, M. (1995). *Mitochondrial plasmids: DNA and RNA.* In: The Molecular biology of plant mitochondria. Editors: Levings, CS III.,Vasil, IK. Kluwer academic publishers, Dordrecht: 61-91.

Brown, GG. (1999). Unique aspects of cytoplasmic male sterility and fertility restoration in *Brassica napus*. *J. Heredity* **90:** 351-356.

Brown, GG., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, WY., Landry, BS. (2003). The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* **35:** 262-272.

Byrd, AD., Schardl, CL., Songlin, PJ., Mogen, Kl., Siegel, MR. (1990). The β-tubulin gene of *Epichloe typhia* from perennial ryegrass (*Lolium perenne* L). *Curr. Genet.* **18:** 347-354.

Cahan, P. and Kennell, JC. (2005). Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. *Mol. Genet. Genomics* **273:** 462-473.

Camlin, MS. (1997). Plant breeding-Achievement and prospects-Grasses. Seeds of Progress. In *Occasional Symposium British Grassland Soc.* Ediors: Weddell, JR. **31:** 2-14.

Castle, ME. and Watson, JN. (1971). A comparison between diploid and tetraploid ryegrass for milk production. *J. Agri. Sci. (Camb)* **77:** 69-76.

Chan, BS., Court, DA., Vierula, PJ., Bertrand, H. (1991). The kalilo linear senescence-inducing plasmid of *Neurospora* is an invertron and encodes DNA and RNA polymerases. *Curr. Genet.* **20:** 225-37.

Chao, S., Sederoff, R., Levings CS III. (1984). Nucleotide sequence and evolution of the 18S ribosomal RNA gene in maize mitochondria. *Nucleic Acids Res.* **12:** 6629-6645.

Chapmann, GP. (1987). The tapetum. *Int. Rev. Cytol.* **107:** 111-125.

Charmet, G., Ravel, C., Balfourier, F. (1997). Phylogenetic analysis in the *Festuca-Lolium* complex using molecular markers and ITS rDNA. *Theor. Appl. Genet.* **94:** 1038-1046.

Chase, C. (2006). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* **23:** 81-90.

Chaudhuri, S., Carrer, H., Maliga, P. (1997). Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J.* **15:** 5958-5964.

Chaumont, F., de Castro Silva Filho, M., Thomas, D., Leterme, S., Boutry, M. (1994). Truncated presequences of mitochondrial F1-ATPase *ß* subunit from *Nicotiana plumbaginifolia* transport CAT and GUS proteins into mitochondria of transgenic tobacco. *Plant Mol. Biol.* **24:** 631-641.

Chaumont, F., Bernier, B., Buxant, R., Williams, ME., Levings CS III., Boutry, M. (1995). Targeting the maize T-*urf13* product into tobacco mitochondria confers methomyl sensitivity to mitochondrial respiration. *Proc. Natl. Acad. Sci. USA* **92:** 1167-1171.

Chaw, SM., Zharkikh, A., Sung, HM., Lau, TC., Li, WH. (1997). Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Mol. Biol. Evol.* **14:** 56-68.

Cho, Y., Adams, KL., Qiu, YL., Kuhlman, P., Vaughn, JC., Palmer, JD. (1998). *A highly invasive group I intron in the mitochondrial coxI gene*. Pp 19-23. In Plant Mitochondria: from gene to function. Editors Moller, I-M., Gardestrom, P., Glimelius, K., Glaser, E. Backhuys Publishers, Leiden, The Netherlands.

Cho, Y., Mower, JP., Qiu, YL, Palmer, JD. (2004). Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. USA* **101:** 17741-17746.

Chung, KR., Leuchtmann, A., Schardl, CL. (1996). Inheritance of mitochondrial DNA and plasmids in the ascomycetes fungus, *Epichloe typhina*. *Genetics* **142:** 259-265.

Clifton, SW., Minx, P., Fauron, MR., Gibson, M., Allen, JO., Sun, H., Thompson, M., Barbazuk, WB., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, RK., Newton, KJ. (2004). Sequence and Comparative Analysis of the Maize NB Mitochondrial Genome. *Plant Physiol.* **136:** 3486-3503.

Collett, M.A., Bradshaw, R.E., Scott, DB. 1995. A mutualistic fungal symbiont of perennial ryegrass contains two different *pyr4* genes, both expressing orotidine-5'-monophosphate decarboxylase. *Gene* **158:** 31-39.

Connolly, V., Ribeiro, MAM., Crowley, JG. (1977). Potential of grass and legume cultivars under Irish conditions. *Proceedings International meeting on animal production from temperate grassland.* Editors: Gilsenan, B. 379-387.

Connolly, V. and Wright-Turner, R. (1984). Induction of cytoplasmic male sterility into ryegrass (*Lolium perenne*). *Theor. Appl. Genet.* **68:** 449-453.

Court, DA. and Bertrand, H. (1992). Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of *Neurospora crassa*. *Curr. Genet.* **22:** 385-397.

Covello, PS. and Gray, MW. (1991). Sequence analysis of wheat mitochondrial transcripts capped in vitro: definitive identification of transcription initiation sites. *Curr. Genet.* **20:** 245-251.

Crane, PR., Friis, EM., Pedersen, KR. (1995). The origin and early diversification of angiosperms. *Nature* **374:** 27-33.

Crespi, M., Vereecke, D., Temmerman, W., van-Montagu, M., Desomer, J. (1994). The fas operon of *Rhodococcus fascians* encodes new genes required for efficient fasciationj of host plants. *J. Bacteriol.* **176:** 2492-2501.

Cui, X., Wise, RP., Schnable, PS. (1996). The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272:** 1334-1336.

de la Canal, L., Crouzillat, D., Flamand, MC., Perrault, A., Boutry, M., Ledoigt, G. (1991). Nucleotide sequence and transcriptional analysis of a mitochondrial plasmid from a cytoplasmic male-sterile line of sunflower. *Theor. Appl. Genet.* **81:** 812-818.

Dabrock, B., Kessler, M., Averhoff, B., Gottschalk, G. (1994). Identification and characterisation of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropyl-benzene and trichlorethene catabolism. *Appl. Environ. Microbiol.* **60:** 853-860.

Daniell, H., Kumar, S., Dufourmantel, N. (2005). Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotech.* **23:** 238-245

Dawson, J., Wilson, ZA., Aarts, MGM., Braithwaite, AF., Briarty, LG., Mulligan, BJ. (1993). Microspore and pollen development in 6-male-sterile mutants of *Arabidopsis thaliana. Can. Journal Bot.-Rev. Can. Bot.* **71:** 629-638.

Debets, F., Yang, X., Griffiths, AJF. (1994). Vegetative incompatibility in *Neurospora*: its effect on horizontal transfer of mitochondrial linear plasmids and senescense in natural populations. *Curr. Genet.* **26:** 113-119.

Dewey, RE., Levings, CS III., Timothy, DH. (1986). Novel recombination in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male sterile cytoplasm. *Cell* **44:** 439-449.

Dewey, RE., Timothy, DH., Levings, CS III. (1987). A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize. *Proc. Natl. Acad. Sci. U S A* **84:** 5374-5378.

Dill, CL., Wise, RP., Schnable, PS. (1997). Rf8 and Rf* mediate unique T-urf13-transcript accumulation, revealing a conserved motif associated with RNA processing and restoration of pollen fertility in T-cytoplasm maize. *Genetics* **147:** 1367-1379.

Dohmen, G. and Tudzynski, PA. (1994). DNA-polymerase-related reading frame (pol-r) in the mtDNA of *Secale cereale*. *Curr. Genet.* **25:** 59-65.

Dufour, E., Mendez, J., Lazaro, JM., de Vega, M., Blanco, L., Salas, M.(2000). An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J. Mol. Biol.* **304:** 289-300.

Duroc, Y., Gaillard, C., Hiard, S., Defrance, MC., Pelletier, G., Budar, F. (2005). Biochemical and functional characterization of ORF138, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in *Brassiceae*. *Biochimie* **87:** 1089-1100.

Duroc, Y., Gaillard, C., Hiard, S., Tinchant, C., Berthome, R., Pelletier, G., Budar, F. (2006) Nuclear expression of a cytoplasmic male sterility gene modifies mitochondrial morphology in yeast and plant cells. *Plant Sci.* **170:** 755-767.

Eggermont, K., Goderis, IJ., Broekaert, WF. (1996). High-throughput RNA extraction from plant samples based on homogenization by reciprical shaking in the presence of a mixture of sand and glass beads. *Plant Mol. Biol. Rep.* **14:** 273-279.

Erickson, L., Kemble, R., Swanson, E. (1989). The *Brassica* mitochondrial plasmid can be sexually transmitted. Pollen transfer of a cytoplasmic genetic element. *Mol. Gen. Genet.* **218:** 419-422.

Escote, LJ., Gabay-Laughnan, SJ., Laughnan, JR. (1985). Cytoplasmic reversion to fertility in CMS-S maize need not involve loss of linear mitochondrial plasmids. *Plasmid* **14:** 264-267.

Farré, JC., Leon, G., Jordana, X., Araya, A. (2001). Cis recognition elements in plant mitochondrion RNA editing. *Mol. Cell Biol.* **21:** 6731-6737.

Fauron, C., Casper, M., Gesteland, R., Albertson, M. (1992). A multi-recombination model for the mtDNA rearrangements seen in maize CMS-T regenerated plants. *Plant J.* **2:** 949-958.

Fauron, CR., Casper, M., Gao, Y., Moore, B. (1995). The maize mitochondrial genome: dynamic, yet functional. *Trends Genet.* **6:** 228-235.

Fei, HM. and Sawhney, VK. (2001). Ultrastructural characterisation of male sterile *33* (ms33) mutant in *Arabidopsis* affected in pollen desiccation and maturation. *Canad. J. Bot.* **79:** 118-129.

Feinberg, AP. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6-13.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783-791.

Fey, J. and Marechal-Drouard, I. (1999). Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria. *Biochem. Biophys. Res. Commun.* **256:** 409-414.

Fey, J., Weil, JH., Tomita, K., Cosset, A., Dietrich, A., Small, I., Marechal-Drouard, L. (2002). Role of editing in plant mitochondrial transfer RNAs. *Gene* **286:** 21-24.

Flavell, R. (1974). A model for the mechanism of cytoplasmic male sterility in plants. with special reference to maize. *Plant Sci. Lett.* **3:** 259-263.

Funk, CR., Murphy J., Huff, DR. (1993). Diversity and vulnerability of perennial ryegrass, tall fescue and Kentuckeys bluegrass. Turfgrass Germplasm Symposium. In Agronomy Abstracts. ASA. p.181. Madison, WI.

Gagliardi, D. and Leaver, CJ. (1999). Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J.* **18:** 3757-3766.

Gagliardi, D. and Gualberto, JM. (2004). Gene Expression in Higher Plant Mitochondria. In *Plant Mitochondria: From Genome to Function.* Editors Day, DA., Millar, H., Whelan, J. **17:** 55-81.

Gallagher, LJ., Betz, SK., Chase, CD. (2002). Mitochondrial RNA editing truncates a chimeric open reading frame associated with S male-sterility in maize. *Curr. Genet.* **42:** 179-184.

Gandolfo, MA., Nixon, KC., Crepet, WL. (1998). A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). *Amer. J. Bot.* **85:** 964-974.

Garcia, JA., Kolacz, K., Studnicka, GM., Gilmore-Herbert, M. (1988). Sequence of integrated S-1 homologous DNA in the normal maize mitochondrial genome. *Nucleic Acids Res.* **16:** 4169-4170.

Gesser-Ulrich, K. and Tudzynski, P. (1992). Transcripts and translation products of a mitochondrial plasmid of *Claviceps purpurea. Curr. Genet.* **21:** 249-254.

Giege, P. and Brennicke, A. (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. USA* **96:** 15324-15329.

Giese H., Lyngkjaer, MF., Stummann, BM., Grell, MN., Christiansen, SK. (2003). Analysis of the structure and inheritance of a linear plasmid from the obligate biotrophic fungus *Blumeria graminis*f.sp.hordei. *Mol. Genet. Genomics* **269:** 699-705.

Girons, IS., Old, IG., Davidson, BE. (1994). Molecular biology of the *Borrelia*, bacteria with linear replicons. *Microbiology* (Reading) **140:** 1803-1816.

Goldberg, RB. Beals, TP., Sanders, PM. (1993), Anther development: Basic principles and practical aaplications. *Plant Cell* **5:**1217-1229.

Griffiths, AJF. (1995). Natural plasmids in filamentous fungi. *Microbiol. Rev.* **59:** 673-685.

Gualberto, JM., Weil, JH., Grienenberger JM. (1990). Editing of the wheat coxIII transcript evidence for tweilve C to U and one U to C conversion and for sequence similarities around editing sites. *Nucleic Acids Res.* **18:** 3771-3776.

Gualberto, JM., Bonnard, G., Lamattina, L., Grienenberger, JM. (1991). Expression of wheat mitochondrial *nad3-rps12* transcription unit: correlation between editing and mRNA maturation. *Plant Cell* **3:** 1109-1120.

Gunge, N., Tamaru, A., Ozawa, F., Sakaguchi, K. (1981). Isolation and characterisation of linear deoxyribonucleic acid plasmids from *Kluyveromyces lactis* and the plasmid-associated killer character. *J. Bacteriol.* **145:** 382-390.

Handa, H., Gualberto, JM., Grienenberger, JM. (1995). Characterisation of the mitochondrial *orfB* gene and its derivative, *orf224,* a chimeric open reading frame specific to one mitochondrial genome of the Polima male-sterile cytoplasm in rapeseed (*Brassica napus* L). *Curr. Genet.* **28:** 546-552.

Handa, H., Itani, K., Sato, H. (2002). Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (*Brassica napus* L.). *Mol. Genet. Genomics* **267:** 797-805.

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. *Nucleic Acids Res.* **31:** 5907-5916.

Hansen, BM. and Marcker, KA (1984). DNA sequence and transcription of a DNA minicircle isolated from male-fertile sugar beet mitochondria. *Nucleic Acids Res.* **12:** 4747-4756.

Hanson, MR. (1991). Plant mitochondrial mutations and male sterility. *Annu. Rev. Genet.* **25:** 461-486.

Hanson, MR. and Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* **16 Suppl:** S154-169.

Hashiba, T., Homma, Y., Hyakumachi, M., Matsuda, I. (1984). Isolation of a DNA plasmid in the fungus *Rhizoctonia solani*. *J. Gen. Microbiol.* **130:** 2067-2070.

He, S., Abad, AR., Gelvin, SR., Mackenzie, SA. (1996). A cytoplasmic male sterility-associated mitochondrial protein causes pollen disruption in transgenic tobacco. *Proc. Natl. Acad. Sci. USA* **93:** 11763-11768.

Heazlewood, JL., Whelan, J., Millar, AH. (2003). The products of the mitochondrial *orf25* and *orfB* genes are FO components in the plant F1FO ATP synthase. *FEBS Lett.* **540:** 201-205.

Hedgcoth, C., el-Shehawi, AM., Wei, P., Clarkson, M., Tarmalis, D. (2002). A chimeric open reading frame associated with cytoplasmic male sterility in alloplasmic wheat with *Triticum timopheevi* mitochondria is present in several *Triticum* and *Aegilops* species, barley, and rye. *Curr. Genet.* **41:** 357-365.

Hedtke, B., Borner, T., Weihe, A. (1997). Mitochondrial and chloroplast phage type RNA polymerases in *Arabidopsis*. *Science* **277:** 809-811.

Hedtke, B., Meixner, M., Gillandt, S., Richter, E., Borner, T., Weihe, A. (1999). Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants *in vivo*. *Plant J.* **17:** 557-561.

Hedtke, B., Borner, T., Weihe, A. (2000). One RNA Polymerase serving two genomes. *EMBO Rep.* **1:** 435-440.

Hermoso, JM., Mendez, E., Soriano, F., Salas, M. (1985). Location of a Serine residue involved in the linkage between the terminal protein and the DNA of the phage phi29. *Nucleic Acids Res.* **13:** 7715-7728.

Hernould, M., Suharsono, S., Litvak, S., Araya, A., Mouras, A. (1993). Male-sterility induction in transgenic tobacco plants with an unedited *atp9* mitochondrial gene from wheat. *Proc. Natl. Acad. Sci. USA* **90:** 2370-2374.

Hernould, M., Suharsono, S., Zabaleta, E., Carde, JP., Litvak, S., Araya, A., Mouras, A. (1998). Impairment of tapetum and mitochondria in engineered male sterile tobacco plants. *Plant Mol. Biol.* **36:** 499-508.

Hesse, M. and Hess, M. (1993). Recent trends in Tapetum research. Cytological and methodological review. *Plant Syst. Evol.* [Suppl]. **7:** 127-145.

Hishinuma, F. and Hirai, K. (1991). Genome organization of the linear plasmid, pSKL, isolated from *Saccharomyces kluyveri. Mol. Gen. Genet.* **226:** 97-106.

Holmes, D. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114:** 193-197.

Horn, R. and Friedt, W. (1999). CMS sources in sunflower: different origin but same mechanism ?. *Theor. Appl. Genet.* **98:** 195-201.

Horsch, R., Fry, J., Hoffmann, N., Eichholtz, D., Rogers, S., Fraley, R. (1985). A simple and general method for transferring genes into plants. *Science* **227:** 1229-1231.

Houliston, GJ. and Olson, MS. (2006). Nonneutral evolution of organelle genes in Silene vulgaris. *Genetics* **174:** 1983-94.

Huff, DR. (1997). RAPD characterization of heterogenous perennial ryegrass cultivars. *Crop. Sci.* **37:** 557-564.

Iwabuchi, M., Kyozuka, J., Shimamoto, K. (1993). Processing followed by complete editing of an altered mitochondrial atp6 RNA restores fertility of cytoplasmic male sterile rice. *EMBO J.* **12:** 1437-1446.

Itana, K. and Handa, H. (1998). Rapeseed mitochondrial ccb206, a gene involved in cytochrome c biogenesis, is co-transcribed with the nad3 and rps12 genes: organization, transcription, and RNA editing of the nad3/rps12/ccb206 locus. *Curr. Genet.* **34:** 318-325.

Izhar, S. and Frankel, R. (1971). Mechanism of male sterility in *Petunia:* the relationship between pH, callase activity in the anthers, and the breakdown of the microsporogenesis. *Theor. Appl. Genet.* **41:** 104-108.

Janska, H., Sarria, R., Woloszynska, M., Arrieta-Montiel, M., Mackenzie, SA. (1998). Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* **10:** 1163-1180.

Jean, M., Brown, GG., Landry, BS. (1997). Genetic mapping of nuclear fertility restorer genes for the 'Polima' cytoplasmic male sterility in canola (*Brassica napus* L.) using DNA markers. *Theor. Appl. Genet.* **95:** 321-328.

Johns, C., Lu, M., Lyznik, A., Mackenzie, S. (1992). A mitochondrial DNA sequence is associated with abnormal pollen development in cytoplasmic male sterile bean plants. *Plant Cell* **4:** 435-449.

Jung, G., Leavitt, MC., Hsieh, J-C., Ito, J. (1987). Bacteriophage PRD1 DNA Polymerase: evolution of DNA polymeraes. *Proc. Natl. Acad. Sci. USA* **84:** 8287-8291.

Kadowaki, KI., Suzuki, T., Kazama, S. (1990). A chimeric gene containing the 5' portion of atp6 is associated with cytoplasmic male sterility of rice. *Mol. Gen. Genet.* **224:** 10-16.

Kalkus, J., Dorrie, C., Fischer, D., Reh, M., Schlegel, HG. (1993). The giant linear plasmid pHG207 from *Rhodococcus sp.* encoding hydrogen autotrophy: characterisation of the plasmid and its termini. *J. Gen. Micrbiol.* **139:** 2055-2065.

Kemble, RJ., Carlson, JE., Erickson, LR., Serynk, JL., Thompson, DJ. (1986). The *Brassica* mitochondrial DNA plasmid and large RNAs are not exclusively associated with cytoplasmic male sterility. *Mol. Gen. Genet.* **205:** 183-185.

Kempken, F., Hermanns, J., Osiewacz, HD. (1992). Evolution of linear plasmids. *J Mol. Evol.* **35:** 502-513.

Kempken, F. (1995). Horizontal transfer of a mitochondrial plasmid. *Mol. Gen. Genet.* **248:** 89-94.

Kesseler, M., Dabbs, ER., Averhoff, B., Gottschalk, G. (1996). Studies on the isopropyl benzene 2,3-dioxygenase and the 3-isopropyl-catechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. *Microbiol.* **142:** 3241-3251.

Kiang, S. (1991). Molecular analysis of Cytoplasmic Male Sterility in ryegrass hybrids. *Ph.D. Thesis.*

Kiang, AS., Connolly, V., McConnell, DJ., Kavanagh, TA. (1993). Cytoplasmic male sterility (CMS) in *Lolium perenne* L.1. Development of a diagnostic probe for the male-sterile cytoplasm. *Theor. Appl. Genet.* **86:** 781-787.

Kiang, AS., Connolly, V., McConnell, DJ., Kavanagh, TA. (1994). Paternal inheritance of mitochondria and chloroplasts in *Festuca pratensis-Lolium perenne* intergeneric hybrids. *Theor. Appl. Genet.* **87:** 681-688.

Kiang, AS. and Kavanagh, TA. (1996). Cytoplasmic male sterility (CMS) in *Lolium perenne* L.2. The mitochondrial genome of a CMS line is rearranged and contains chimeric atp9 gene. *Theor. Appl. Genet.* **92:** 308-315.

Kim, EK., Jeong, JH., Youn, HS., Koo, YB., Roe, JH. (2000). The terminal protein of a linear mitochondrial plasmid is encoded in the N-terminus of the DNA polymerase gene in white-rot fungus *Pleurotus ostreatus*. *Curr. Genet.* **38:** 283-290.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111-120.

Kinashi, H., Shimaji, M., Sakai, A. (1987). Giant linear plasmids in Streptomyces which code for antibiotic biosynthesis genes. *Nature* **328:** 454-456.

Klein, M., Eckert-Ossenkopp, U., Schmiedeberg, I., Brandt, P., Unseld, M., Brennicke, A., Schuster, W. (1994). Physical mapping of the mitochondrial genome of Arabidopsis thaliana by cosmid and YAC clones. *Plant J.* **6:** 447-455.

Klein, RR., Klein, PE., Mullet, JE., Minx, P., Rooney, WL., Schertz, KF. (2005). Fertility restorer locus Rf1 [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor. Appl. Genet.* **111:** 994-1012.

Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S., Ullrich, H., Brennicke, A. (1996). copia-, gypsy- and LINE-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. *Genetics* **142:** 579-585.

Knoop, V. (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.* **46:** 123-139.

Kohler, RH., Horn, R., Lossl, A., Zetsche, K. (1991). Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the atpA gene. *Mol. Gen. Genet.* **227:** 369-376.

Koizuka, N., Imai, R., Iwabuchi, M., Sakai, T., Imamura, J. (2000). Genetic analysis of fertility restoration and accumulation of ORF125 mitochondrial protein in the Kosena radish (*Raphanus sativus* L. cv. Kosena) and a *Brassica napus* restorer line. *Theor. Appl. Genet.* **100:** 949-955.

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Muraase, J., Sakai, T., Kawasaki, S., Imamura, J. (2003). Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male sterile Kosena radish. *Plant J.* **34:** 407-415.

Kopecky, D., Lukaszewski, AJ., Gibeault, V. (2005). Reduction of Ploidy level by Androgenesis in Intergeneric *Lolium-Festuca* hybrids for Turf Grass breeding. *Crop Sci.* **45:** 274-281.

Kotera, E., Tasaka, M., Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*. **433:** 326-330.

Koultunow, AM., Treuttner, J., Cox, KH., Wallroth, M., Goldberg, RB. (1990). Different temporal and spatial gene expression patterns occur during anther developmemt. *Plant Cell* **2**: 1201-1224.

Ku, S., Yoon, H., Suh, HS., Chung, YY. (2003). Male sterility of thermosensitive genic male sterile rice is associated with premature programmed cell death of the tapetum. *Planta* **217:** 559-565.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., Mikami, T. (2000). The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Res.* **28:** 2571-2576.

Kuchuk, N., Sytnyk, K., Vasylenko, M., Shakhovsky, A., Komarnytsky, I., Kushnir, S., Gleba, Y. (2006). Genetic transformation of plastids of different *Solanaceae* species using tobacco cells as organelle hosts. *Theor. Appl. Genet.* **113:** 519-527.

Kuhn, K., Bohne, AV., Liere, K., Borner, T (2007). *Arabidopsis* phage-type RNA polymerases: accurate in vitro transcription of organellar genes. *Plant Cell* **19:** 959-971.

Kumar, S., Tamura, K., Jakobsen, IB., Nei, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis Software. *Bioinformatics* **17:** 1244-1245.

Landgren, M., Zetterstrand, M., Sundberg, E., Glimelius, K. (1996). Alloplasmic male-sterile Brassica lines containing *B. tournefortii* mitochondria express an ORF 3' of the *atp6* gene and a 32 kDa protein. *Plant Mol. Biol.* **32:** 879-890.

Laroche, J., Peng, Li., Bousquet, J. (1995). Mitochondrial DNA and Monocot-Dicot Divergence Time. *Mol. Biol. Evol.* **12:** 1151-1156.

L'Homme, Y. and Brown, GG. (1993). Organizational differences between cytoplasmic male sterile and male fertile *Brassica* mitochondrial genomes are confined to a single transposed locus. *Nucleic Acids Res.* **21:** 1903-1909.

L'Homme, Y., Stahl, RJ., Li, XQ., Hameed, A., Brown, GG. (1997). *Brassica* nap cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the pol CMS-associated *orf224* gene. *Curr. Genet.* **31:** 325-335.

Latch, GCM., Christensen, MJ., Samuels, GJ. (1984). Five endophytes of *Lolium* and *Festuca* in New Zealand. *Mycotaxon* **20:** 535-550.

Leino, M., Teixeira, R., Landgren, M., Glimelius, K. (2003). *Brassica napus* lines with rearranged *Arabidopsis* mitochondria display CMS and a range of developmental aberrations. *Theor. Appl. Genet.* **106:** 1156-1163.

Levings, CS III. and Sederoff, RR. (1983). Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize. *Proc. Natl. Acad. Sci. USA* **80:** 4055-4059.

Levings, CS III. (1990). The Texas cytoplasm of maize: Cytoplasmic. male sterility and disease susceptibility. *Science* **250:** 942-947.

Levings, CS III. (1993). Thoughts on Cytoplasmic Male Sterility in cms-T Maize. Plant *Cell* **5:** 1285-1290

Li, WH., Wu, C.-l., Luo, C.-C., (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. *Mol Biol. Evol.* **2:** 150-174.

Li, WH. and Graur, D. (1991). Fundamentals of molecular evolution. Sinaur Associates, Sunderland, MA.

Li, XQ., Jean, M., Landry, BS., Brown, GG. (1998). Restorer genes for different forms of brassica cytoplasmic male sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. *Proc. Natl. Acad. Sci. USA* **95:** 10032-10037.

Lilly, JW., Bartoszewski, G., Malepszy, S., Harvey, MJ. (2001). A major deletion in the cucumber mitochondrial genome sorts with the MSC phenotype. *Curr. Genet.* **40:** 144-151.

Lu, B., Wilson, RK., Phreaner, CG,, Mulligan, RM., Hanson, MR. (1996). Protein polymorphism generated by differential RNA editing of a plant mitochondrial rps12 gene. *Mol. Cell Biol.* **16:** 1543-1549.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, ML., Mireau, H., Peeters, N., Renou, JP., Szurek, B., Taconnat, L., Small, I. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16:** 2089-2103.

McAllistair, WT. and Raskin, CA. (1993). The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol. Microbiol.* **10:** 1-6.

McCormick, S. (1993). Male gametophyte development. *Plant Cell* **5**: 1265-1275.

Malek, O., Lattig, K., Brennicke, A., Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* **15:** 1403-1411.

Maleszka, R., Skelly, PJ., Clark-Walker, GD. (1991). Rolling circle replication of DNA in yeast mitochondria. *EMBO J.* **10:** 3923-3929.

Mandel, M. and Higra, A. (1970). Calcium dependant bacteriophage DNA infection. *J. Mol. Biol.* **53:** 154.

Manthey, GM. and McEwen, JE. (1995). The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COXI locus of *Saccharomyces cerevisiae*. *EMBO J.* **14:** 4031-4043.

Marcinko-Keuhn, M., Yang, X., Debets, F., Jacobson, DJ., Griffiths, AJF. (1994). A kalilo like linear plasmid in Louisiana field isolates of the pseudohomothallic fungus *Neurospora tetrasperma*. *Curr. Genet.* **26:** 336-343.

Marechal-Drouard, L., Cosset, A., Remacle, C., Ramamonjisoa, D., Dietrich, A. (1996). A single editing event is a prerequisite for efficient processing of potato mitochondrial phenylalanine tRNA. *Mol. Biol. Evol.* **16:** 3504-3510.

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J., Goldberg, RB. (1990). Induction of male sterility in plants by a chimeric ribonuclease gene. *Nature* **347:** 737-741.

Martin, W., Gierl, A., Saedler, H. (1989). Molecular evidence for pre-Cretaceous angiosperm origin. *Nature* **339:** 46-48.

Mascarenhas, JP. (1990). Gene Activity during pollen development. *Annu. Rev. Physiol. Mol. Biol.* **41:** 317-338.

Matsuoka, Y., Yamazaki, Y., Ogihara, Y., Tsunewaki, K. (2002). Whole Chloroplast Genome Comparison of Rice, Maize and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Mol. Biol. Evol.* **19:** 2084-2091.

Meijer, WJ., Horcajadas, JA., Salas, M. (2001). *phi29* Family of Phages. *Microbiol. Mol. Biol. Rev.* **65:** 261-287.

Meinhardt, F., Kempken, F., Kamper, J., Esser, K. (1990). Linear plasmids among eukaryotes: fundamentals and application. *Curr. Genet.* **17:** 89-95.

Miyamoto, T., Obokata, J., Sugiura, M. (2004). A site-specific factor interacts directly with its cognate RNA editing site in chloroplast transcripts. *Proc. Natl. Acad. Sci. USA* **101:** 48-52.

Miyashita, S., Hirochika, H., Ikeda, JE., Hashiba, T. (1990). Linear plasmid DNAs of the plant pathogenic fungus *Rhizoctonia solani* with unique terminal structures. *Mol. Gen. Genet.* **220:** 165-171.

Mogen, KL., Siegel, MR., Schardl, CL. (1991). Linear DNA plasmids of the perennial ryegrass choke pathogen, *Epichloe typhina* (Clavicipitaceae). *Curr. Genet.* **20:** 519-526.

Moneger, F., Smart, CJ., Leaver, CJ. (1994). Cell-specific regulation of gene expression in mitochondria during anther development in sunflower. *Plant Cell* **6:** 811-825.

Mori, Y., Kimura, S., Saotome, A., Kasai, N., Sakaguchi, N., Uchiyama, Y., Ishibashi, I., Yamamoto, I., Chiku, H., Sakaguchi, K. (2005). Plastid DNA polymerases from higher plants, *Arabidopsis thaliana. Biochem. Biophys. Res. Comm.* **334:** 43-50.

Mower, JP. and Palmer, JD. (2006). Patterns of partial RNA editing in mitochondrial genes of *B. vulgaris. Mol. Genet. Genomics* **276:** 285-293.

Mower, JP., Touzet, P., Gummow, S., Delph, LF., Palmer, JD. (2007). Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol.* **9:** 135.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol. Plant* **15:** 473-479.

Murphy, G. and Kavanagh, TA. (1988). Speeding up the sequencing of double stranded DNA. *Nucleic Acids Res.* **16:** 5198.

Nakai, S., Noda, D., Kondo, M., Terachi, T. (1995). High-level expression of a mitochondrial orf522 gene from the male sterile sunflower is lethal to *E.coli. Breed. Sci.* **45:** 233-236.

Nakai, R., Kikuo, S., Kurosawa, S., Hiroshiro, S. (2000). The Basidomycete fungus *Flammulina velutipes* harbors two linear mitochondrial plasmids encoding DNA and RNA Polymerases. *FEMS Microbiol. Lett.* **190:** 99-102.

Neuwirt, J., Takenaka, M., van der Merwe, JA., Brennicke, A. (2005). An in vitro RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. *RNA* **11:** 1563-1570.

Newton, KJ., Mariano, JM., Gibson, CM., Kuzmin, E., Gabay-Laughnan, S. (1996). Involvement of S2 episomal sequences in the generation of NCS4 deletion mutation in maize mitochondria. *Dev. Genet.* **19:** 277-286.

Newton, KJ., Gabay-Laughnan, S., De Paepe R. (2004). Mitochondrial mutations in plants. In *Plant Mitochondria: From Genome to Function.* Editors: Day, DA., Millar, AH., Whelan, J. **17:** 122-141. Kluwer.

Nivison, HT., Sutton, CA., Wilson RK., Hanson, MR. (1994). Sequencing, processing, and localization of the petunia CMS-associated mitochondrial protein. *Plant J.* **5:** 613-623.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., Kadowaki, K. (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics.* **268:** 434-445.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223:** 1-7.

Oeser, B. and Tudzynski, P. (1989). The linear mitochondrial plasmid pClK1 of the phytopathogenic fungus *Claviceps purpurea* may code for a DNA polymerase and an RNA polymerase. *Mol. Gen. Genet.* **217:** 132-140.

Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S., Tsunewaki, K., (2005). Structural dynamics of cereal mitochondrial genomes as revealed by complete

nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* **33:** 6235-6250.

Ono, Y., Sakai, A., Takechi, K., Takio, S., Takusagawa, M., Takano, H. (2007). NtPolI-like1 and NtPolI-like2, bacterial DNA polymerase I homologues isolated from BY-2 cultured tobacco cells, encode DNA polymerases engaged in DNA replication in both plastids and mitochondria. *Plant Cell Physiol.* **48:** 1679-1692.

Oro, A., Newton, KJ., Walbot, V. (1985). Molecular analysis of the inheritance and stability of the mitochondrial genome of an inbred line of maize. *Theor. Appl. Genet.* **70:** 287-293.

Owen, HA. and Makaroff, CA. (1995). Ultra structure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L.) Heynh ecotype Wassilewskija (*Brassicaceae*). *Protoplasma* **185:** 7-21.

Pacini, E., Franchi, GG., Hesse, M. (1985). The tapetum: its form, function, and possible phylogeny in Embryophyta. *Plant Syst. Evol.* **149:** 155-185.

Paillard, M., Sederoff, RR., Levings, CS III. (1985). Nucleotide sequence of the S-1 mitochondrial DNA from the cytoplasm of maize. *EMBO J.* **4:** 1125-1128.

Palmer, JD. and Zamir, D. (1982). Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. *Proc. Natl. Acad. Sci. USA* **79:** 5006-5010.

Palmer, JD., Shields, CR., Cohen, DB., Orton, TJ. (1983). An unusual mitochondrial DNA plasmid in the genus *Brassica*. *Nature* **301:** 725-728.

Palmer, JD. and Shields, CR. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* **307:** 411-415.

Palmer, JD. and Herbon, LA. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28:** 87-97.

Palmer, RG., Albertsen, MC., Horner, HT., Skorupska, H. (1992). Male sterility in soybean and maize: developmental comparisons. *Nucleus* **35:** 1-18.

Palmer, JD., Adams, KL., Cho, Y., Parkinson, CL., Qiu, YL., Song, K. (2000). Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci. USA* **97:** 6960-6966.

Papini, A., Mosti, S., Brighigna, L. (1999). Programmed cell death events during tapetum development of angiosperms. *Protoplasma* **207:** 213-221.

Parkinson, CL., Mower, JP., Qiu, YL., Shirk, AJ., Song, K., Young, ND., DePamphilis, CW., Palmer, JD. (2005). Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evol. Biol.* **5:** 73-84.

Paxson-Sowders, DM., Dodrill, CH., Owen, HA., Makaroff, CA. (2001). DEX1 a novel plant protein is required for exine pattern development during pollen development in *Arabidopsis*. *Plant Physiol.* **127:** 1739-1749.

Pelletier, G. and Budar, F. (2007). The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. *Curr. Opin. Biotech.* **18:** 121-125.

Phreaner, CG., Williams, MA., Mulligan, RM. (1996). Incomplete editing of *rps12* transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. *Plant Cell* **8:** 107-117.

Piffanelli, RI. and Murphy, DJ. (1998) Novel organelles and targeting mechanisms in the anther tapetum. *Trends Plant Sci.* **3:** 250-253.

Presier PR, Wilson, RJM., Moore, PW., McCreedy, S., Hajibagheri, MAN., Blight, KJ., Strath, M., Williamson, DH. (1996). Recombination associated with replication of malarial mitochondrial DNA. *EMBO J.* **15:** 684-693.

Pring, DR. and Lonsdale, DM. (1985) Molecular biology of higher plant mitochondrial DNA. *Int. Rev. Cytol.* **97:** 1-46.

Pring, DR. and Van Tang, H. (2001). Mitochondrial *atp6* transcript editing during microgametogenesis in male-sterile sorghum. *Curr. Genet.* **39:** 371-376.

Prioli, LM., Huang, J., Levings, CS III. (1993). The plant mitochondrial open reading frame *orf221* encodes a membrane-bound protein. *Plant Mol. Biol.* **23:** 287-295.

Pruitt, KD. and Hanson, MR. (1991) Transcription of the Petunia mitochondrial CMS-associated *pcf* locus in male sterile and fertility restored lines. *Mol. Gen. Genet.* **227:** 348-355.

Pryer, KM., Schneider, H., Zimmer, EA. (2002). Deciding among green plants for whole genome studies. *Trends Plant Sci.* **7:** 550-554.

Qiu, YL. and Palmer, JD. (1999) Phylogeny of early land plants: insights from genes and genomes. *Trends Plant Sci.* **4:** 26-30.

Regan, SM. and Moffatt, BA. (1990). Cytochemical analysis of pollen development in wild type *Arabidopsis* and a male sterile mutant. *Plant Cell* **2:** 877-889.

Rhee, SY. and Sommerville, CR. (1998). Tetrad pollen formation in quartet mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall. *Plant J.* **15:** 79-88.

Robison, MM. and Horgen, PA. (1998). Widespread distribution of low copy number variants of mitochondrial plasmid plasmid pEM in the genus *Agaricus*. *Fungal Genet. Biol.* **26:** 62-70.

Robison, MM. and Wolyn, DJ. (2002). Comple organization of the mitochondrial genome of the petaloid carrot. *Mol. Gen. Genomics* **268:** 232-239.

Robison, MM. and Wolyn, DJ. (2005). A mitochondrial plasmid and plasmid- like RNA and DNA polymerases encoded within the mitochondrial genome of carrot (*Daucus carota* L.). *Curr. Genet.* **47:** 57-66.

Rogers, HJ. (2006). Programmed cell death in floral organs: how and why do flowers die? *Ann. Bot. (Lond).* **97:** 309-315.

Rohe, M., Schrunder, J., Tudzynski, P., Meinhardt, F. (1992). Phylogenetic relationships of linear, protein-primed replicating genomes. *Curr. Genet.* **21:** 173-176.

Roque, E., Gomez, M., Ellul, P., Wallbraun, M., Madueno, F., Beltran, JP., Canas, LA. (2007). The PsEND1 promoter: a novel tool to produce genetically engineered male-sterile plants by early anther ablation. *Plant Cell Rep.* **26:** 313-25.

Ruiz, ON. and Daniell, H. (2005). Engineering cytoplasmic male sterility via the chloroplast genome by expression of {beta}-ketothiolase. *Plant Physiol.* **138:**1232-1246.

Ryohsuke, N., Kikuo, S., Shin-ichi, K., Hiroshiro, S. (2000). Basidiomycetous fungus *Flammulina velutipes* harbors two linear mitochondrial plasmids encoding DNA and RNA polymerases. *FEMS Microbiol. Letts.* **190:** 99-102.

Sabar, M., Gagliardi, D., Balk, J., Leaver, CJ. (2003) ORFB is a subunit of F1F(O)-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Rep.* **4:** 381-386.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406-425.

Sakai, A., Suzuki, T., Nagata, N., Sasaki, N., Miyazawa, Y., Saito, C., Inada, N., Nishimura, Y., Kuroiwa, T. (1999). Comparative analysis of DNA synthesis activity in plastid-nuclei and mitochondrial-nuclei simultaneously isolated from cultured tobacco cells. *Plant Sci.* **140:** 9-19.

Sakamoto, W., Kondo, H., Murata, M., Motoyoshi, F. (1996). Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by chloroplast mutator. *Plant Cell* **8:** 1377-1390.

Salas, M. (1991). Protein priming of DNA replication. *Annu. Rev. Biochem.* **60:** 39-71.

Sambrook, J., Fritsch, E., Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* 2nd edition (Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press).

Sandhu, APS., Abdelnoor, RV., Mackenzie, SA. (2007). Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. *Proc. Natl. Acad. Sci. USA* **104:** 1766-1770.

Sarria, R., Lyznik, A., Vallejos, CE., MacKenzie, SA. (1998). A cytoplasmic male sterility-associated mitochondrial peptide in common bean is post-translationally regulated. *Plant Cell* **10:** 1217-1228.

Satoh, M., Kubo, T., Nishizawa, S., Estiati, A., Itchoda, N., Mikami, T. (2004). The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs. *Mol. Genet. Genomics* **272:** 247-256.

Saumitou-Laprade, P., Pannenbecker, G., Maggouta, F., Jean, R., Michaelis, G (1989). A linear 10.4 kb plasmid in the mitochondria of *Beta maritima*. *Curr. Genet.* **16:** 181-186.

Schardl, CL., Pring, DR., Lonsdale, DM. (1985). Mitochondrial DNA rearrangements associated with fertile revertants of S-type male-sterile maize. *Cell* **43:** 361-368.

Schardl, CL., Leuchtmann, A., Tsai, HF., Collett, MA., Watt, DM., Scott, DB. (1994). Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*. *Genetics* **136:** 1307-1317.

Schardl, CL., Leuchtmann, A., Spiering, MJ. (2004) Symbiosis of grasses with seedborne fungal endophytes. *Annu. Rev. Plant Biol.* **55:** 315-340.

Schuster, W., Wissinger, B., Unseld, M., Brennicke, A. (1990). Transcripts of the NADH-dehydrogenase subunit 3 gene are differentially edited in Oenothera mitochondria. *EMBO J.* **9:** 263-269.

Schuster, W., Ternes, R., Knoop, V., Hiesel, R., Wissinger, B., Brennicke, A. (1991). Distribution of RNA editing sites in Oenothera mitochondrial mRNAs and rRNAs. *Curr. Genet.* **20:** 397-404.

Schnable, PS. and Wise, RP. (1998). The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci.* **3**: 175-180.

Sena, EP., Revet, B., Moustacchi, E. (1986). In *vivo* homologous recombination intermediates of yeast mitochondrial DNA analyzed by electron microscopy. *Mol. Gen. Genet.* **202:** 421-428.

Seurinck, J., Truettner, J., Goldberg, RB. (1990). The nucleotide sequence of an anther-specific gene. *Nucleic Acids Res.* **18:** 3403.

Shikanai, T. and Yamada, Y. (1988). Properties of the circular plasmid-like DNA, B4, from mitochondria of cytoplasmic male-sterile rice. *Curr. Genet.* **13:** 441-443.

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol. Life Sci.* **63:** 698-708.

Shinada, T., Kikuchi, Y., Fujimoto, R., Kishitani, S. (2006). An alloplasmic male-sterile line of Brassica oleracea harboring the mitochondria from *Diplotaxis muralis* expresses a novel chimeric open reading frame, orf72. *Plant Cell Physiol.* **47:** 549-553.

Shukla, AK., Vijayaraghavan, MR., Chaudhry, B. (1998). *Biology of Pollen*. New Delhi: APH Publishing Corporation.

Singh, M. and Brown, GG. (1991). Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell* **3:** 1349-1362.

Small, ID., Suffolk, R., Leaver, CJ. (1989). Evolution of plant mitochondrial genomes via sub-stoichiometric intermediates. *Cell* **58:** 69-76.

Smart, CJ., Moneger F., Leaver, CJ. (1994). Cell-Specific Regulation of Gene Expression in Mitochondria during Anther Development in Sunflower. *Plant Cell* **6:** 811-825.

Smith, AG. and Pring, DR. (1987). Nucleotide sequence and molecular characterization of a maize mitochondrial plasmid-like DNA. *Curr. Genet.* **12:** 617-623.

Song, J. and Hedgcoth, C. (1994). A chimeric gene (*orf256*) is expressed as protein only in cytoplasmic male sterile lines of wheat. *Plant Mol. Biol.* **26:** 535-539.

Steiglitz, H. (1977). Role of β-1,3-glucanase in postmeiotic microspore release. *Dev. Biol.* **57:** 87-97.

Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A., Sugiura, M. (2005). The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genomics*. **272:** 603-615.

Suzuki, K., Takeda, H., Tsukaguchi, T., Egawa, Y. (2001). Ultrastructutal study on degeneration of tapetum in anther of snap bean (*Phaesolus vulgaris* L) under heat stress. *Sex Plant Reprod.* **13:** 293-299.

Takenaka, M,, Neuwirt, J., Brennicke, A. (2004). Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* **32:** 4137-4144.

Tang, HV., Pring, DR., Shaw, LC., Salazar, RA., Muza, FR., Yan, B., Schertz, KF. (1996). Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. *Plant J.* **10:** 123-133.

Tang, HV., Chen, W., Pring, DR. (1999). Mitochondrial *orf107* transcription, editing, and nucleolytic cleavage conferred by the gene *Rf3* are expressed in sorghum pollen. *Sex Plant Reprod.* **12:** 53-59.

Teixeira, RT., Farbos, I., Glimelius, K. (2005a). Expression levels of meristem identity and homeotic genes are modified by nuclear-mitochondrial interactions in alloplasmic male-sterile lines of *Brassica napus*. *Plant J.* **42:** 731-742.

Teixeira, RT., Knorpp, G., Glimelius, K. (2005b). Modified sucrose, starch, and ATP levels in two alloplasmic male-sterile lines of *B. napus*. *J. Exp. Bot.* **56:** 1245-1253.

Thomas, H. and Humphreys, MO. (1991). Progress and potential of interspecific hybrids of *Lolium* and *Festuca*. *J. Agric. Sci. (Camb.)* **117:** 1-8.

Thomas, HM., Morgan, WG., Humphreys, MW. (2003). Designing grasses with a future-combining the attributes of *Lolium* and *Festuca*. *Euphytica* **133:** 19-26.

Thompson, JD., Higgins, DG., Gibson, TJ. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic acids Res.* **22:** 4673-4680.

Torri, AF. and Englund, PT. (1995). A DNA polymerase beta in the mitochondrion of the trypanosomatid *Crithidia fasciculata*. *J. Biol. Chem.* **270:** 3495-3497.

Tsuchiya, T., Toriyama, K., Yoshikawa, M., Ejiri, S., Hinata, K. (1995). Tapetum-specific expression of the gene for an endo-beta-1,3-glucanase causes male sterility in transgenic tobacco. *Plant Cell Physiol.* **36:** 487-494.

Turpen, T., Garger, SJ., Marks, MD., Grill, LK. (1987). Molecular cloning and physical characterisation of a *Brassica* linear mitochondrial plasmid. *Mol. Gen. Genet.* **209:** 227-233.

Unseld, M., Marienfeld, JR., Brandt, P., Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15:** 57-61.

Vaughn, JC., Mason, MT., Sper-Whitis, GL., Kuhl-man, P., Palmer, JD. (1995). Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of *Peperomia. J. Mol. Evol.* **41:** 563-572.

Vickery, DB. and Griffiths, AJ. (1993). Transcription of the kalilo linear senescence plasmid from *Neurospora intermedia. Plasmid* **29:** 180-192.

Wahleithner, JA. and Wolstenholme, DR. (1987). Mitochondrial plasmid DNAs of broad bean: nucleotide sequences, complex secondary structures, and transcription. *Curr. Genet.* **12:** 55-67.

Wang, M., Hoekstra, S., van Bergen, S., Lamers, GE., Oppedijk, BJ., van der Heijden, MW., de Priester, W., Schilperoort, RA. (1999). Apoptosis in developing , anthers and the role of ABA in this process during androgenesis in *Hordeum vulgare* L. *Plant Mol. Biol.* **39:** 489-501.

Wang, A., Xia, Q., Xie ,W., Dalta, R., Selvara, G. (2003). The Classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc. Natl. Acad. Sci. USA* **100:** 14487-14492.

Wang, J., Machado, C., Pannaccione, DG., Tsai, HF., Schardl, CL. (2004). The determinant step in ergot alkaloid biosynthesis by an endophyte of perennial ryegrass. *Fungal Genet. Biol.* **41:** 189-198.

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y., Liu, YG. (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* **18:** 676-687.

Ward, CG. (1995). The Texas male-sterile cytoplasm of maize. In *The Molecular Biology of Plant Mitochondria.* Editors: Levings, CS III. and Vasil, IK. **3:** 433-459. Kluwer Academic.

Warmke, HE. and Lee, SJ. (1977). Mitochondrial degeneration in Texas cytoplasmic male sterile corn anthers. *J. Heredity* **68:** 213-222.

Warmke, HE. and Lee, SJ. (1978). Pollen abortion in T cytoplasmic male-sterile corn (*Zea mays*): A suggested mechanism. *Science* **200:** 561-563.

Weber, B., Borner, T., Weihe, A. (1995). Remnants of a DNA polymerase gene in the mitochondrial DNA of *Marchantia polymorpha. Curr. Genet.* **27:** 488-490.

Wen, L. and Chase, CD. (1999). Pleiotropic effects of a nuclear restorer-of-fertility locus on mitochondrial transcripts in male-fertile and S male-sterile maize. *Curr. Genet.* **35:** 521-526.

Wintz, H., Chen, HC., Sutton, CA., Conley, CA., Cobb, A. (1995). Expression of the CMS-associated *urfS* sequence in transgenic petunia and tobacco. *Plant Mol. Biol.* **28:** 83-92.

Wise, RP., Pring, DR., Gengenbach, BG. (1987). Mutation to the male fertility and toxin insensitivity in Texas (T)-cytoplasm maize is associated with a frameshift in a mitochondrial open reading frame. *Proc. Natl. Acad. Sci. USA* **84:** 2858-2862.

Wise, RP., Dill, CL., Schnable, PS. (1996). Mutator-induced mutations of the rf1 nuclear fertility restorer of T-cytoplasm maize alter the accumulation of T-urf13 mitochondrial transcripts. *Genetics* **143:** 1383-1394.

Wise, RP., Gobelman-Werner, K., Pei, D., Dill, CL., Schnable, PS. (1999). Mitochondrial transcript processing and restoration of male fertility in T-cytoplasm maize. *J. Heredity* **90:** 380-385.

Wit, F. (1974). Cytoplasmic male sterility in ryegrasses (*Lolium* spp.) detected after intergeneric hybridization. *Euphytica* **23:** 31-38.

Wolfe, KH., Li, WH., Sharp, PM. (1987). Rate of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84:** 9054-9058.

Wolfe, KH., Guoy, M., Yang, YW., Sharp, PH., Li, WH. (1989). Date of monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. USA* **86:** 6201-6205.

Worrall, D., Hird, DL., Hodge, R., Paul, WW., Draper, J., Scott, R. (1992). Premature dissolution of the microsprorocyte callose wall causes male sterility in transgenic tobacco. *Plant Cell* **4:** 759-771.

Worsham, PL. and Bolen, PL. (1990). Killer toxin production in *Pichia acaciae* is associated with linear DNA plasmids. *Curr. Genet.* **18:** 77-80.

Wu, X. and Roy, KL. (1993). Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterisation of its RNA transcripts. *J. Bacteriol.* **175:** 37-52.

Wu, SSH., Platt, KA., Ratnayake, C., Wang, TW., Ting, JTL., Huang, AHC. (1997). Isolation and characterisation of neutral lipid containing organelles and globuli filled plastids from *Brassica napus* tapetum. *Proc. Natl. Acad. Sci. USA* **94:** 12711-12716.

Wu, HM. and Cheung, AY. (2000). Programmed cell death in plant reproduction. *Plant Mol. Biol.* **44:** 267-281.

Xu, F., Morin, C., Mitchell, G., Ackerley, C., Robinson, BH. (2004). The role of the *LRPPRC* (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome *c* oxidase) I and COX III mRNA. *J. Biochem.* **382:** 331-336.

Yang, YW., Lai, KN., Tai, PY., Li, WH. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *J. Mol. Evol.* **48:** 597-604.

Young, EG. and Hanson, MR. (1987). A fused mitochondrial gene associated with cytoplasmic male sterility is developmentally regulated. *Cell* **50:** 41-49.

Zabala, G., O'Brien-Vedder, C., Walbot, V. (1987). S2 episome of maize mitochondria encodes a 130-kilodalton protein found in male sterile and fertile plants. *Proc. Natl. Acad. Sci. USA* **84:** 7861-7865.

Zabala, G. and Walbot, V. (1988). An S1 episomal gene of maize mitochondria is expressed in male sterile and fertile plants of the S-type cytoplasm. *Mol. Gen. Genet.* **211:** 386-392.

Zabala, G., Gabay-Laughnan, S., Laughnan, JR. (1997). The nuclear gene *Rf3* affect the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. *Genetics* **147:** 847-860.

Zabaleta, E., Mouras, A., Hernould, M., Suharsono, S., Araya, A. (1996). Transgenic male-sterile plant induced by an unedited *atp9* gene is restored to fertility by inhibiting its expression with antisense RNA. *Proc. Natl. Acad. Sci. USA* **93:** 1259-1263.

Zwierzykowski, Z., Tayyar, R., Brunell, M., Lukaszewski, AJ. (1998). Genome recombination in intergeneric hybrids between tetraploid *Festuca pratensis* and *Lolium multiflorum*. *J. Heredity* **89:** 324-328.