# HOMOGENISATION IN THE RIBOSOMAL RNA GENES OF AN *EPICHLOË* ENDOPHYTE HYBRID

A thesis presented in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy
in
Molecular Genetics

at Massey University, Palmerston North
New Zealand

Austen Rawlinson Deldyck Ganley

1999

*For Denis and Phil Ganley*

# ABSTRACT

Homogenisation mechanisms in the ribosomal RNA genes were investigated using *Epichloë* fungal endophyte interspecific hybrid isolate, Lp1. The two progenitor isolates, *Neotyphodium lolii* Lp5 and *E. typhina* E8, were used for comparison. Three areas of homogenisation were examined.

The first area involved characterisation of extraordinary length heterogeneity in the rDNA of Lp1. This was shown by Southern analyses on single-spore isolates to be present intragenomically and localised to the intergenic spacer (IGS). Length heterogeneity is not a feature of either progenitor, suggesting it is a consequence of the hybridisation. The length heterogeneity was shown to result from copy number variation of sub-repeats in the IGS, which is consistent with unequal crossing over occurring in the rDNA, suggesting that unequal crossing over plays a role in homogenisation. Multi-variant repeat PCR mapping of the sub-repeat array revealed that the ends of the array behave differently, and biased initiation of recombination is discussed. Several results are not consistent with homogenisation by unequal crossing over and the potential roles of gene conversion and extrachromosomal rDNA circles in homogenisation are discussed. Finally, evidence is presented that suggests the rate of homogenisation is very rapid. A group I intron is present in the 28S *rrn* gene of Lp1, and is widespread in the *Epichloë* endophytes. Closely-related introns in other fungal 18S *rrn* genes provide evidence for intron transposition.

The second area involved testing the hypothesis that the presence of one type of rDNA sequence in Lp1 is the result of interlocus homogenisation. CHEF gel electrophoresis revealed that Lp1 and Lp5 have at least five rDNA arrays organised as major and minor loci, an unusual situation in fungi. The organisation in E8 could not be determined. One potential rDNA-DNA junction was cloned but has not been analysed.

The final area initially involved testing the hypothesis that interlocus homogenisation of 5S rRNA gene arrays occurs more slowly than that of rDNA arrays in hybrids. However the 5S rRNA genes in the *Epichloë* endophytes were shown to be organised as dispersed copies, not in tandem arrays. Shared polymorphisms between Lp1, Lp5 and E8 may indicate the homogenisation rate of these dispersed repeats is slower, and gene conversion as a homogenisation mechanism is discussed. The 5S rRNA genes are located on the same chromosomal bands as the rDNA in Lp1 and Lp5, and therefore are markers that demonstrate the rDNA-containing Lp5 chromosomes are present in Lp1. This and the CHEF results provide evidence for interlocus homogenisation of the rDNA having occurred in Lp1, and extends observation of this phenomenon to fungi.

# ACKNOWLEDGEMENTS

During my years at Massey University (now a decade!) I have come to know many wonderful people who have contributed to this thesis in a myriad of ways. Somehow I am meant to, and want to, distil into a few words all your encouragement, ideas, distractions, patience, forbearance, beers, inspiration, etc., etc. - a nigh-on impossible task! So to everyone who I omit to mention personally - cheers, I'll buy you a beer sometime!

First I want to thank my parents, Janine and Garth, for their unconditional love and support throughout. It was them who instilled in me the belief that I could achieve the goals that I set, whatever they were, and it is that which sees me where I am today. To them and the rest of my family - Beccy, Jules and Karen, Denise and Steve, and the many others, thanks for everything (I couldn't have done it without you guys) - especially for understanding when I have been focussed on work to the exclusion of all else.

Carolyn and David - what can I say? If helping out at all times of the day and night wasn't enough, you then let me live with you, drink your beer, and have parties when you're not there! And that's not even counting work or putting the thesis together... Carolyn, I fail to understand how people in other labs can operate properly without you there - does that express how much I owe you for all your help in the lab? You are a rock. Really. And David - the world would be a poorer place without beer and mushrooms, and who better to enjoy both with.

Rich, Mike and Hale, it was you guys who influenced my impressions on what science and scientists were about and (oddly enough) helped me decide to do this "thing". Our many adventures in places far-and-wide have kept me going throughout adversity, and provided a kind of "reality grounding" for which I am very grateful to have. I hope there will be more!

Over the years I have lived with some wonderful people, who, I suspect, have shaped my life in more ways than I realise. Tania, my longest and dearest flatmate! Living with you was constant fun, and I miss not having you around. Thanks for everything! Justin - you narrowly scrape into this section by virtue of around nine months flatting, but the same couldn't be said as far as friendship is concerned. A Justin in the hand is worth two in the bush, I always say, and have yet to be proved wrong! With Jaki and Lucy, and Steve and Dave I had the unlikely pleasure of having two perfect flats - you guys are great! I also want to thank Bruce and Toshi for letting me take over their place - it was a perfect place for writing. And cheers for the various harvests too!

Anna, I enjoyed living with you when we were together. I guess things change and are not always as we may think or like. I hope we have both learned from our experiences.

I would like to thank many dear friends for helping me see life outside of scientific world. To Damian, Paul and Carolyn, Maria, Cybele, Francesca, Carmel and Tony, Kirsty, Jenny, Jules and Morgan - viva la revolucion!! To Linda, Kate, Peng and Ina, Marty, Mark and Vanessa, Sue, God, Blair, Mike, Marcus, Ian, Roo and Blake I owe thanks for many-and-varied "extra curricular" activities and liberties. On the political front people like the Summer of Resistance rabble, the Waihopai "hippy-type individuals", Edwina, Patrick, Eleven, Ian, Stewart, Chris, Maryanne and Dion, Kevin, Gerard, and Don are a constant source of inspiration. Thanks for your dedication and doing all the stuff that I couldn't do - kia kaha!

I would also like to thank many dear friends in the scientific world. It is people such as Marion, Miranda, Mick and Linda, Paul, Rebecca, Rochelle, Rose, Martin, Ant, Pete, Steve and Basil who make this a great place to work! I would especially like to thank Brendon for all the advice, beers and just plain chatting we shared.

The many people who have passed through the Scott Base lab and surrounds during my time here have all made a mark in their own ways. I would like to thank in particular Christina for our many discussions on endophytes, Lisa for saving me from crises, Emily for beers and rugby, Sirinda, Jo, Bek and Seth. Thanks also go to David Penny for helpful advice and ideas on the evolutionary side of things, Mike Christensen for helpful advice and ideas on the endophyte side of things, and Max Scott as my co-supervisor for trying to tether my flights-of-fancy back to scientific reality!

I would like to thank Massey University for a Doctoral Scholarship that partially supported me financially through this PhD.

And for the dessert course! My most sincere thanks and gratitude go to my supervisor, Barry Scott. I know that the course of study I chose to pursue was not one in your immediate area of interest, nor was one that you would have preferred me to do. But since I made the decision you have provided unstinting input and support of the highest quality, and for your willingness to become familiar with an obscure and at times confusing field to an extent that few would match I am extremely grateful. I have thoroughly enjoyed my time working with you, and I think we got the mix of independence and tutelage about right. I hope this thesis reflects that!

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER ONE

# INTRODUCTION

## 1.1 BIOLOGY OF THE *EPICHLOË* ENDOPHYTES

The *Epichloë* grass endophytes, which form the basis of this study, are a group of ascomycetous filamentous fungi from the family *Clavicipitaceae*, which includes the fungus responsible for St Anthony's Fire (ergotism) from contaminated rye. These endophytes form symbiotic relationships with pasture grasses of the sub-family *Pooideae* (Clay, 1990). They produce a range of secondary alkaloids which help protect the plant from the effects of insect and mammalian herbivory, drought and nematodes, as well as giving greater field persistence (Clay, 1990; Siegal and Schardl, 1991). The alkaloids include a tremorgenic mycotoxin responsible for the neurotoxic disorder in grazing mammals, ryegrass staggers, which is of economic importance to New Zealand (Scott and Schardl, 1993).

The *Epichloë* endophytes exhibit two types of life cycle, and this is related to their method of reproduction (Schardl, 1996). The sexual forms, which are designated as *Epichloë* species, exhibit pathogenic symptoms through production of external stromata on the inflorescences of their hosts. Stromal production, which represents the sexual stage of the fungus, prevents maturation of the host inflorescences and leads to sterility - a phenomenon known as "choking". The asexual forms are designated as *Neotyphodium* species (previously *Acremonium*; Glenn *et al.*, 1996). These are entirely endophytic and are disseminated via the host grass seed, never resulting in choke. The strict seed transmission of these asexual endophytes means they can be seen in a sense as maternal genetic elements of the host grass. Molecular phylogenetic studies indicate that the anamorphic (asexual) *Neotyphodium* endophytes have evolved from the teleomorphic (sexual) *Epichloë* endophytes (Bucheli and Leutchmann, 1996; Schardl *et al.*, 1991). The division in life cycles is not strict. There are some sexual *Epichloë* species that are also seed transmitted and do not always produce choke. Hence these *Epichloë* endophytes form a continuum of associations with their host grasses, ranging from antagonism (complete formation of choke) through pleiotropy (partial formation of choke and partial seed transmission) to mutualism (strict seed transmission), based on the relative importance of the sexual cycle (Clay, 1988; Schardl, 1996).

### 1.1.1 INTERSPECIFIC HYBRIDISATION OF THE *EPICHLOË* ENDOPHYTES

Recent studies have shown that hybridisation between species of the *Epichloë* endophytes has played a major role in their evolution, with several independent hybridisation events described (Moon *et al.*, 1999; Schardl *et al.*, 1994; Tsai *et al.*, 1994). In fact, all the *Neotyphodium* endophytes that have been characterised to date, except *N. lolii* (LpTG-1; see Section 1.1.2), are interspecific hybrids (Moon *et al.*, 1999). In some cases there appear to have been successive hybridisation events, producing multiple hybrid endophytes (e.g. *Neotyphodium coenophialum*; Tsai *et al.*, 1994). The propensity of these endophytes to hybridise may be a means of overcoming Muller's ratchet - the progressive accumulation of deleterious mutations in asexual

species (Muller, 1964). Hybridisation is in contrast to many asexual species which maintain a haploid genome. This is thought to aid a population in long-term survival by making the effect of mutations more acute (Lynch *et al.*, 1993). The frequent hybridisations between these *Epichloë* endophytes may be an alternative to maintaining a high selection level as a means of escaping Muller's ratchet. Alternatively hybridisation may be a feature of these endophytes that allows increased genetic redundancy giving increased evolutionary plasticity, or may be a result of the host plant disarming the antagonistic choke (Collett, 1994).

The mechanism by which these endophytes hybridise is not known, and it is interesting that repeated attempts to construct hybrids artificially have all been unsuccessful (M.J. Christensen and C.L. Schardl, personal communication). Thus even though interspecific hybridisation is an important phenomenon in evolutionary time, it seems to be a very rare event in biological time, although it may be that appropriate conditions for hybridisation have not been found yet. The most likely scenario for hybridisation involves co-infection of a single grass plant with two endophytes, via an ascospore germinating on an already-infected plant. These endophytes would then undergo anastomosis (hyphal fusion) followed by karyogamy (nuclear fusion) to create a hybrid (Tsai *et al.*, 1994). This scenario is feasible as the *Epichloë* endophytes are vegetatively compatible with each other, even between different mating populations (Chung and Schardl, 1997).

### 1.1.2 THE ORIGIN OF *EPICHLOË* ENDOPHYTE ISOLATE, LP1

This study is based on work performed on an asexual *Neotyphodium* endophyte isolate called Lp1. Lp1 is endophytic in *Lolium perenne* (perennial ryegrass), and is one of two *Neotyphodium* taxonomic groupings found in *L. perenne*. The taxonomic grouping that Lp1 falls into is known as *L. perenne* taxonomic grouping 2 (LpTG-2), with the other known as LpTG-1 or *N. lolii* (Christensen *et al.*, 1993). Molecular phylogenetic studies using genetic loci, isozyme data and microsatellite data concluded that Lp1 was an interspecific hybrid, with the progenitors being an LpTG-1 (*N. lolii*) isolate and an isolate from the sexual species *Epichloë typhina*, which also inhabits *L. perenne* (Moon *et al.*, 1999; Schardl *et al.*, 1994). The only locus that does not indicate a hybrid origin is the ITS region of the rDNA (see Section 1.2.1). Isolate E8 from *E. typhina* appears to correspond to the *E. typhina* portion of the Lp1 genome. An extant isolate of LpTG-1 corresponding to the LpTG-1 portion of the Lp1 genome has not been found, but isolate Lp5 is very close (Schardl *et al.*, 1994). Therefore throughout this study I have used LpTG-1 isolate Lp5 and *E. typhina* isolate E8 as the closest progenitors of Lp1, and they are referred to as the progenitors. An independent LpTG-2 isolate, Lp2, has been found which I have also used in this study. Molecular and morphological data indicate that Lp1 and Lp2 are very similar (Schardl *et al.*, 1994; M. Christensen and A.R.D. Ganley, unpublished results).

## 1.2 RIBOSOMAL RNA GENES

### 1.2.1 THE 18S - 5.8S - 28S RIBOSOMAL RNA GENES

The 18S, 5.8S and 28S ribosomal RNA genes (*rrn* genes) are organised as a cistron that is repeated many times in most eukaryotes (including fungi; Long and Dawid, 1980). These genes encode RNA species (rRNA) that form the ribosome, which is responsible for translating mRNA into polypeptide. The cistrons are arranged in a head-to-tail tandem array with internal transcribed spacers (ITS) separating the *rrn* genes within a unit, and an intergenic spacer (IGS) separating adjoining units. The entirety of the DNA coding for these genes is known as the rDNA, and this consists of one or more of these arrays of *rrn* genes and spacers. Each *rrn* gene/spacer array is also referred to as a locus. It is presumed that multiple copies of the *rrn* genes are required to produce sufficient rRNA for ribosome biosynthesis, although in most organisms there is an excess of rDNA units and only a subset are transcribed, a phenomenon known as redundancy (Hadjiolov, 1984; Long and Dawid, 1980). The rDNA is transcribed by RNA polymerase I, and this occurs in a structure specific to the rDNA - the nucleolus. The nucleolus is visible in the nucleus with appropriate staining under the microscope and contains the rDNA that is being transcribed. It has been well-studied (Hadjiolov, 1984) and there are a number of proteins that are localised to it, however it is not a membrane-bound organelle. In species for which the chromosomes can be seen under the light microscope, the rDNA clusters are visible as constrictions in the chromosome (s). These rDNA clusters have been called nucleolus organiser regions (NOR) because a nucleolus forms around them. Although the nucleolus is primarily thought of in terms of its role in ribosomal RNA transcription and ribosome biogenesis, it is becoming clear that it also plays roles in other, seemingly unrelated, cellular processes (see Garcia and Pillus, 1999).

The *rrn* genes show a remarkable amount of sequence conservation across a wide range of species, whilst the spacer elements diverge more rapidly (Hillis and Dixon, 1991). It is widely held that the difference in conservation between the genes and the spacers is an outcome of the genes being under functional constraints whilst the spacers are relatively free from such constraints. The high level of conservation found in the *rrn* genes is not constant throughout the genes. Within the genes, especially the 28S *rrn* gene, there are regions that are very highly conserved and others that are relatively free to diverge in sequence. These differences are presumably the result of different regions of the gene being under different functional constraints (Gerbi, 1986). The pattern of sequence conservation in the *rrn* genes and spacers is mirrored by the pattern of length conservation. The varying rates of sequence divergence have been exploited for phylogenetic analyses, as appropriate regions of the rDNA can be selected for comparisons at different phylogenetic levels. This has made the rDNA a popular choice in phylogenetic studies (Hillis and Dixon, 1991).

One source of length variation in the *rrn* genes that does not appear to relate to functional constraint is the presence of insertion sequences. These can take two forms. The first are insertion sequences that interrupt the coding region and inactivate the gene - indeed there is evidence that transcription of these interrupted genes is suppressed (Hadjiolov, 1984). The best-characterised examples of this kind of insertion sequence are those found in *Drosophila melanogaster*, and they have been shown to resemble non long terminal repeat retrotransposons (Jakubczak *et al.*, 1990; Pellegrini *et al.*, 1977). Obviously a high proportion of these insertion sequences in the rDNA would be deleterious to the cell, so they only occupy a subset of the rDNA units. The second type of insertion sequence found in the rDNA are group I introns. These also interrupt the coding region, but they splice out of the RNA transcript (usually via a self-splicing reaction; Cech, 1990), recreating a functional gene (hence the epithet "intron"). Self-splicing of a group I intron was the first definitive example of RNA having catalytic function (Kruger *et al.*, 1982). Their ability to splice out of a gene means they can be present in every rDNA unit without affecting ribosomal function, and several encode an endonuclease that aids their spread throughout the rDNA units.

Length variation of the rDNA unit is mainly the result of IGS length variation, and the IGS can vary widely in size between species (see Appels and Honeycutt, 1986). It is normally maintained at a constant length in the rDNA within a species, however in some species the length fluctuates within populations, individuals, and even arrays. The IGS sequence from a number of species has been determined, and more have been restriction mapped. Despite significant sequence divergence even between closely related species, many IGS are found to contain sub-repeats - small, tandemly-repetitive sequences within the IGS (Appels and Honeycutt, 1986). In many species these repetitive elements are duplications of the rDNA promoter and some act as enhancers of transcription (Moss *et al.*, 1985). Therefore the number of sub-repeats in such systems is an important regulator of the transcriptional activity of the rDNA. In a number of instances where length variation of the IGS within or between individuals has been investigated, it results from variation in the number of these IGS sub-repeats (see Section 4.1 for citations). Length variation of the IGS has been used to distinguish closely related species or as an intraspecific population marker (Appels and Honeycutt, 1986).

### 1.2.2 THE RIBOSOMAL DNA IN LP1

An rDNA unit from Lp1 was cloned using phage λ library screening (C.A. Young, unpublished results). The λ clone containing an entire rDNA unit plus some of the flanking rDNA units is designated λPN1. λPN1 was mapped (Gan, 1992) using the restriction endonucleases *Sal*I and *Eco*RI, and the map and its relationship to the structural organisation of an rDNA unit is shown in Figure 1.1. From the λ clone it can be seen that *Sal*I divides an rDNA unit into a 5.6 kb fragment containing primarily the three *rrn* genes, and a 4.1 kb fragment containing primarily the IGS. *Eco*RI divides an rDNA unit into a 3.2 kb fragment containing primarily the 28S *rrn* gene, and a fragment containing primarily the other *rrn* genes and the IGS. This second *Eco*RI

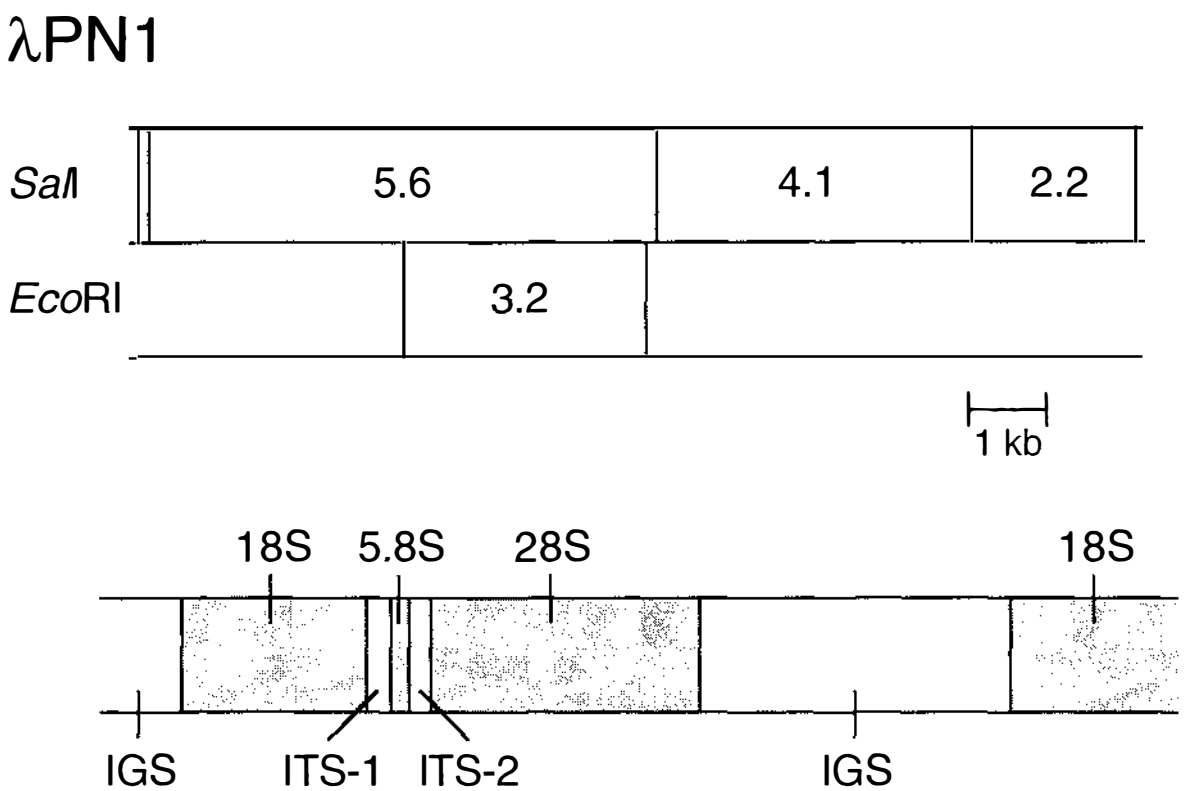



Figure 1.1 Restriction map of λPN1

The restriction enzyme map of the Lp1 ribosomal DNA-containing λ clone, λPN1, is shown in the upper part of the figure. This gives the positions of the *Sal*I and *Eco*RI sites in the λ clone. Sizes are in kb, and the 5.6 kb and 4.1 kb *Sal*I fragments from λPN1 that were cloned into pUC118 to generate pPN49 and pPN50 respectively (Table 2.1) are shown. The lower part of the figure shows the organisation of the ribosomal RNA genes and spacers in λPN1. The shaded regions are the ribosomal RNA genes (18S, 5.8S and 28S *rrn* genes) and the white regions are the intergenic spacer (IGS) and internal transcribed spacers (ITS-1 and ITS-2), as indicated. A complete rDNA unit can be defined by the two *Sal*I sites, one cleaving just before the 5' end of the 18S *rrn* gene and the other cleaving in the 3' end of the 28S *rrn* gene. *Eco*RI divides an rDNA unit into two fragments. One includes primarily the 5.8S and 28S *rrn* genes, and the other includes primarily the IGS and 18S *rrn* gene which is not complete in this clone. The 2.2 kb *Sal*I fragment is a partial copy of the 5.6 kb *Sal*I fragment that is interrupted by the *Sal*I site from the λEMBL3A multicloning site. The left-most *Sal*I site is the other *Sal*I site from the λEMBL3A multicloning site.

fragment is not complete in λPN1, but its expected size (determined from the other bands in λPN1) is 6.5 kb. The two full-length *Sal*I fragments from λPN1 (5.6 kb and 4.1 kb) were subcloned into pUC118 (K. Marriott, unpublished results and Gan, 1992; Table 2.1). When a *Sal*I Southern blot of Lp1 was probed with the 4.1 kb *Sal*I fragment from λPN1, there was evidence of length heterogeneity in this region (D. Watt, unpublished results; Ganley, 1993).

### 1.2.3 THE 5S RIBOSOMAL RNA

The 5S ribosomal RNA (5S rRNA) genes are also present in multiple copies in the genome. They are usually organised as tandem arrays like the rDNA, but are normally not associated with the rDNA (Hadjiolov, 1984). In some cases the 5S rRNA genes are organised in the same tandem array as the *rrn* genes, but this linkage is thought to be a derived characteristic (Drouin and Moniz de Sá, 1995). In a few cases the 5S rRNA genes are not tandemly arranged, but are present as individual, dispersed copies (see Section 4.10 for citations). The 5S rRNA genes are transcribed by RNA polymerase III and are all very similar lengths (around 120 bp; Singer and Berg, 1991). The primary transcriptional regulatory regions are located within the gene itself, as has been found for other RNA polymerase III-transcribed genes (Geiduschek and Tocchini-Valentini, 1988). Like the *rrn* genes, the 5S rRNA gene has regions of high and low sequence conservation, presumably as a result of varying functional constraints.

## 1.3 CONCERTED EVOLUTION

### 1.3.1 PATTERNS OF CONCERTED EVOLUTION

Concerted evolution (Zimmer *et al.*, 1980) is the term used to describe the unusual evolutionary behaviour of multigene families whose genes show a great deal of similarity to each other within an array and within a species but accumulate differences between species. This was first demonstrated experimentally in the rDNA in *Xenopus* by Brown *et al.* (1972), and concerted evolution has continued to be most thoroughly studied in the rDNA. This ability of individual repeats in a multigene family to evolve in concert rather than independently is believed to result from a process that is able to convert all the repeats in an array into a single type, a process known as homogenisation (Arnheim, 1983; Dover, 1982; Dover *et al.*, 1982). Although concerted evolution is primarily discussed in reference to eukaryotes, there is evidence for homogenisation occurring in prokaryotes as well (reviewed in Liao, 1999).

This homogenisation (leading to the observed pattern of concerted evolution) is believed to be important in maintaining the functionality of multigene families. A selectionist argument for the evolution of homogenisation would be along the lines that as there is redundancy in the number of rDNA units in a genome (Hadjiolov, 1984), mutations in *rrn* genes that render the gene non-functional will have no phenotypic effect until a large number of genes in the array are mutated.

This would lead to degeneration of the rDNA. Homogenisation of the array means that selection is able to act on the rDNA as if it was a single locus - if a functional *rrn* gene is spread through the array this will survive, but if a non-functional gene is spread this will be selected against (Williams, 1990). Therefore homogenisation is seen as a means of protecting the integrity of multigene families. Another argument for homogenisation (the "genome turnover" argument) suggests that it occurs naturally as a result of the turnover processes (such as recombination) that are found throughout the genome. This explains why homogenisation occurs in non-coding repetitive DNA, and perhaps how this repetitive DNA arises in the first place (Charlesworth *et al.*, 1986; Smith, 1976). In either case homogenisation is a process which counteracts the effects of mutation, and therefore the level of variation in a multigene family is the product of the competing processes of homogenisation and mutation. Implicit here is the idea that, when looking at a single multigene array, the degree of variability observed is largely the outcome of homogenisation, not of selection (Dover, 1989) - selection operates at the array level.

While homogenisation has been inferred from the observed patterns of similarities and differences in multigene families, it has only been directly demonstrated twice; in the rDNA of hybrids in lizards (Hillis *et al.*, 1991) and cotton (Wendel *et al.*, 1995). The concept of concerted evolution resulting from homogenisation of repeat arrays in multigene families is widely accepted, however the precise *modus operandi* of the homogenising mechanism(s) is not well understood (for recent reviews see Elder and Turner, 1995; Li, 1997; Liao, 1999).

Comparison between species reveals high levels of conservation in the *rrn* genes, but highly divergent spacer regions, as discussed above. However homogenisation is believed to act at the DNA level, and therefore be blind to the difference between genes and spacers (Kellogg and Appels, 1995). Furthermore, it is believed to be a stochastic process - any sequence is able to spread by homogenisation until it reaches a point where selection may play a role. How then can a homogenisation process that acts indiscriminately on genes and spacers result in drastically different patterns of evolution for these regions? The difference arises because the spacers can essentially "hitch-hike" (Begun and Aquadro, 1992) with the more functionally-constrained *rrn* genes during the homogenisation process - that is, as a repeat unit spreads throughout an array a wide range of spacer variants will be functional, but only a very small number of *rrn* gene variants will be tolerated. Therefore a particular unit containing functional *rrn* genes that spreads through an array could be associated by chance with a wide range of different spacer variants (Kellogg and Appels, 1995; Smith, 1973).

### 1.3.2 MECHANISMS OF HOMOGENISATION

The history of research into the mechanisms of homogenisation has primarily involved hypotheses for a mechanism being proposed on theoretical grounds, followed by searches for experimental evidence supporting or rejecting the mechanism. Currently the two mechanisms most commonly invoked as responsible for homogenisation are gene conversion (Birky and

Skavaril, 1976; Edelman and Gally, 1970; Nagylaki and Petes, 1982; Ohta, 1977) and unequal crossing over (Ohta, 1976; Perelson and Bell, 1977; Smith, 1973). These mechanisms, which fall under the general term "recombination", are proposed to drive a single repeat unit within an array to fixation in that array, with selection presumably acting to bring about removal of unfit genes that spread to a high proportion.

Unequal crossing over is a process that occurs specifically with repeated elements. In the case of a tandemly-repeated array, unequal crossing over occurs when two arrays (from sister chromatids or separate chromosomes) align and undergo a reciprocal exchange. The process is unequal if the arrays do not align perfectly in register with one another - i.e. if repeat number *k* from one array does not align with the equivalent repeat number *k* on the other array. The outcome of such a reciprocal unequal exchange is that one array will have more repeat units and the other less repeat units than they had before the exchange. The extent of repeat number change is dependent on the degree of misalignment. Therefore the outcome of an unequal cross over event is perturbation of the ratios of repeat variants that were initially present in the arrays. Smith (1973) showed that successive unequal cross over events, initiated at random in the array, can lead to one variant in the array eventually being fixed throughout the array, aided by genetic drift. The process of turnover leading to fixation of one repeat type is termed homogenisation. Smith's classic diagrammatical representation of this process is shown in Figure 1.2.

Gene conversion involves the unidirectional transfer of information from one DNA duplex to another, and so is not in this sense a reciprocal event. Gene conversion is much more intuitive as a homogenisation mechanism (as well as being easier to model mathematically; Ohta, 1983) because its very action "corrects" the sequence of a stretch of recipient DNA to that of the donor DNA - an explicitly homogenising action. Like unequal crossing over it relies on pairing of homologous stretches of DNA and therefore can occur between repetitive elements. It can occur between sister chromatids and different chromosomes, and also may occur within an array. Gene conversion was first discovered in fungi (reviewed in Whitehouse, 1982), and much of the work exploring the molecular details of gene conversion has been done in fungi.

Experimental evidence for both unequal crossing over and gene conversion has been found in the rDNA (Gangloff *et al.*, 1996; Muscarella and Vogt, 1993; Petes, 1980; Szostak and Wu, 1980), but the relative roles of these in the homogenisation process is uncertain. Resolution of this debate has been hampered by difficulties in distinguishing these mechanisms experimentally with such a large number of essentially identical genes. Also, unequal crossing over and gene conversion are believed to be mechanistically linked (Holliday, 1964; Lichten and Goldman, 1995; Meselson and Radding, 1975; Szostak *et al.*, 1983), perhaps even being different outcomes of the same intermediate, as gene conversion and crossing over are often associated (Klein, 1995). However the isolation of mutants affecting either gene conversion or unequal crossing over demonstrates that they are under some degree of independent control (Aguilera and Klein, 1988; Gysler-Junker *et al.*, 1991; Orr-Weaver and Szostak, 1985; Whitehouse, 1982). In

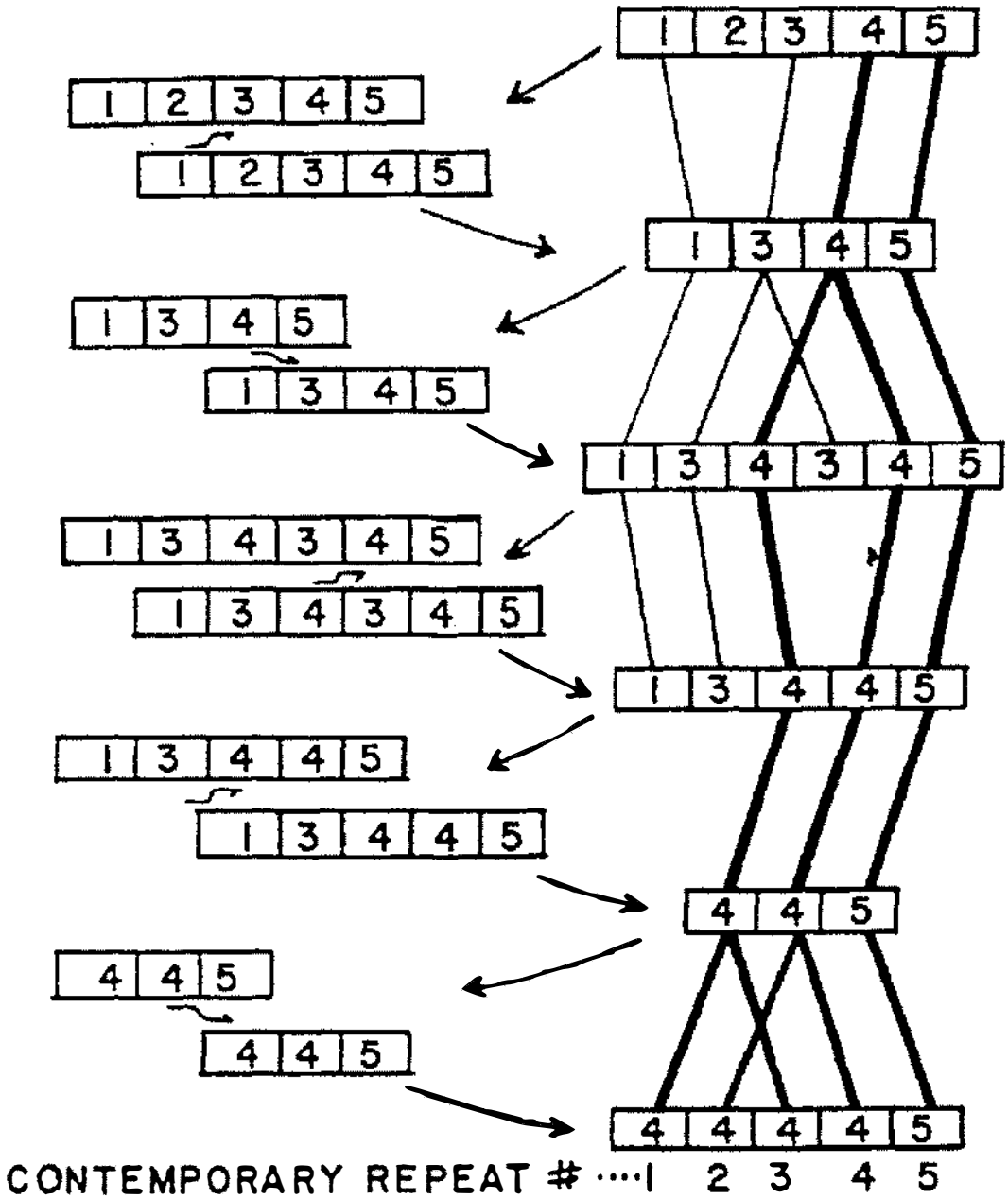



Figure 1.2 Unequal crossing over as a mechanism for homogenisation of a tandem array

Schematic representation showing how repeated cycles of unequal crossing over can lead to fixation of one repeat type in the array. Different numbers in the boxes represent different repeat variants. Unequal crossing over results in variation of repeat number and of repeat variants in the array. In each case recombination occurs between two like arrays (e.g. between sister chromatids). Only one of the two products of the cross over makes it through to the next round - drift. In reality the effect of drift will be much smaller. Origins of the repeat variants in the "contemporary" array are indicated by heavy lines. Diagram taken from Smith (1973).

part these problems in distinguishing crossing over from gene conversion are also the result of uncertainty in the precise molecular details of these processes. Despite considerable effort over a number of years, the molecular mechanism (s) of recombination in eukaryotes remains largely elusive (e.g. Haber 1999). Several models of recombination have been proposed (Holliday, 1964; Meselson and Radding, 1975; Szostak *et al.*, 1983), and problems in determining which of these is the *bone fide* mechanism may be a result of more than one mechanism being employed. The precise molecular details of recombination are not critical for a broad understanding of how these mechanisms can bring about homogenisation, with many of the proposed models equally able to operate to bring about homogenisation. However in some cases making detailed predictions of the outcome of one or other mechanism for homogenisation requires knowledge of the specific molecular details by which they operate.

Recombination can be meiotic or mitotic, and these types also appear to be under at least partially independent control (reviewed in Klein, 1995; Orr-Weaver and Szostak, 1985). As mentioned above, recombination events can occur between different substrates: between homologous chromosomes (allelic or classical recombination); between repeats on non-homologous chromosomes (interchromosomal recombination); between sister chromatids (sister chromatid recombination) and even between repeats on the same chromatid (intrachromatid recombination) - the latter two processes together are known as intrachromosomal recombination (adapted from Jinks-Robertson and Petes, 1993).

The preceding discussion of homogenisation has focussed on homogenisation of a single array. In many cases multigene families are present on more than one array per genome. In these cases homogenisation is extended to homogenisation occurring within an array and homogenisation occurring between different arrays (Ohta and Dover, 1983). There is much experimental evidence to suggest that the rate of homogenisation within an array (intralocus homogenisation) is greater than that occurring between arrays (interlocus homogenisation; e.g. Appels and Dvorák, 1982; Appels and Honeycutt, 1986; Arnheim *et al.*, 1982; Coen and Dover, 1983b; Krystal *et al.*, 1981; Liao *et al.*, 1997; Nagylaki, 1984; Polans *et al.*, 1986; Saghai-Maroof *et al.*, 1984; Schlötterer and Tautz, 1994). Polanco *et al.* (1998) have presented results that suggest the rDNA IGS undergoes relatively rapid interlocus homogenisation in *D. melanogaster*, but the ITS is homogenised primarily intrachromosomally. They propose that this is the result of the exchange of IGS sub-repeats as they undergo meiotic X-Y pairing. This indicates that a variety of homogenisation mechanisms may be active and that the unit of homogenisation is important in the patterns that are generated. That interlocus homogenisation occurs is not disputed - the two examples where homogenisation has been directly demonstrated both involve interlocus homogenisation occurring between different rDNA arrays that have been introduced into the same genome through hybridisation (Hillis *et al.*, 1991; Wendel *et al.*, 1995), and there are other examples where it seems likely to have occurred (van Houten *et al.*, 1993; Zhang and Sang, 1999). Although in many cases there are differences between arrays, these are usually less than the differences between species, indicating that interlocus homogenisation is still more rapid than

speciation. There have been some reports of persistent differences between arrays (Carranza *et al.*, 1996; O'Donnell and Cigelnik, 1997; O'Kane *et al.*, 1996; Vogler and DeSalle, 1994; Zijlstra *et al.*, 1995), but the significance of these is not clear. Studies by Hijri *et al.* (1999) and Hosny *et al.* (1999) show that multiple rDNA types in the fungus *Scutellospora castanea* are the result of independent nuclei being maintained in the fungus that have different rDNA sequences, and this indicates that care must be taken in assessing the long-term persistence of rDNA variants.

In the case of the cotton hybrid where rDNA interlocus homogenisation had occurred, there was not a corresponding interlocus homogenisation of the 5S rRNA genes (Cronn *et al.*, 1996). This and other results reported in the literature suggested that different forms of the 5S rRNA genes are maintained in the genome more readily than different forms of the rDNA (Dvorák, 1989; Kellogg and Appels, 1995; Reddy and Appels, 1989; Sastri *et al.*, 1992). On the basis of these results it appears that interlocus homogenisation of the 5S rRNA gene arrays occurs much more slowly than interlocus homogenisation of the rDNA arrays.

## 1.4 AIMS AND OBJECTIVES OF THIS STUDY

The general aim of this study was to investigate the ribosomal RNA genes to attempt to obtain evidence for or against various homogenisation mechanisms. The hybrid *Neotyphodium* endophyte, Lp1, was chosen for this investigation as an rDNA unit had been cloned, length heterogeneity in the rDNA had been discovered and the progenitors of the hybrid were fairly well characterised. In addition, Lp1 had been shown to be a composite of its two progenitors with respect to genetic loci examined, isozyme data and microsatellite data. The one exception was the ITS region of the rDNA, which appeared to be solely derived from one progenitor. The reason (s) for this discrepancy was not determined, but I hypothesised that it was the result of interlocus homogenisation, as has been found for other hybrids. Various reports in the literature led me to propose the hypothesis that interlocus homogenisation of 5S rRNA gene arrays occurred much more slowly than interlocus homogenisation of rDNA arrays.

Thus the initial objectives of this study were:

1. To characterise the length heterogeneity in the rDNA of Lp1, and compare this to the situation in the progenitors in order to determine the origin of the length heterogeneity and any association it may have with homogenisation of the rDNA.

2. To determine the reason (s) for only one type of rDNA ITS sequence found in the hybrid, Lp1, and whether this has any association with interlocus homogenisation of the rDNA.

3. To determine whether there were 5S rRNA gene arrays from both progenitors in the hybrid, Lp1, and to compare the homogenisation of the 5S rRNA genes with that of the rDNA.

# CHAPTER TWO

# MATERIALS AND METHODS

# 2.1 STRAINS AND GROWTH CONDITIONS

## 2.1.1 FUNGAL AND BACTERIAL STRAINS, λ CLONES AND PLASMIDS

Fungal and bacterial strains, λ clones and plasmids used in this study are listed in Table 2.1.

## 2.1.2 FUNGAL GROWTH CONDITIONS

Fungal isolates were cultured on potato dextrose agar (PD agar; Section 2.2.1) plates at 22°C for two to five weeks, depending on the growth rate of the isolate. Subculturing of fungal cultures was performed by cutting out a small mycelial block and inoculating this onto a fresh PD agar plate.

When mycelia was required for DNA extractions (Section 2.3.1) or protoplasting (Section 2.19.1), a mycelial block approximately 8 x 8 mm was ground to a fine suspension in 500 µl of PD broth (Section 2.2.1). This was inoculated into a flask containing 30 ml of PD broth and incubated at 22°C shaking at 250 rpm until a thick mycelial suspension was achieved (four to seven days).

## 2.1.3 SINGLE-SPORE FUNGAL PURIFICATION

Spore suspensions of fungal isolates were prepared by cutting a 10 x 10 mm mycelial block and agitating this in 150 µl of sterile water. This spore suspension was spread onto a PD agar (2% w/v agar) plate and incubated at 22°C for 48 h. The plates were examined under a dissecting microscope, germinating single spores were cut out under the dissecting microscope and patched to individual PD agar (Section 2.2.1) plates. Manipulations of fungal cultures (this Section and Section 2.1.2) were performed in a laminar flow workstation (Gelman Sciences).

## 2.1.4 BACTERIAL GROWTH CONDITIONS

*Escherichia coli* cultures were grown at 37°C for 16 h on LB agar plates or in LB broth (Section 2.2.2) shaking at 200 rpm.

# 2.2 MEDIA AND COMMON SOLUTIONS

All media and solutions were prepared with MilliQ water. All media was autoclaved at 121°C for 15 min.

Table 2.1 Fungal and Bacterial Strains, λ Clones and Plasmids

| Strain, λ clone or plasmid | Relevant characteristics | Source or reference |
|---|---|---|
| Fungal strains: | | |
| *Neotyphodium* sp. (LpTG-2)[a] | | |
| Lp1O | Original Lp1 isolate | F.R. Murray |
| Lp1A0 | Laboratory culture derived from Lp1O | A.R.D. Ganley |
| Lp1C0 | Laboratory culture derived from Lp1O | M.J. Christensen |
| Lp1D0 | Laboratory culture derived from Lp1O | M.A. Collett |
| Lp1F0 | Laboratory culture derived from Lp1O | R.D. Johnson |
| Lp2G0 | Original Lp2 isolate | M.J. Christensen |
| Lp1A3 | Single-spore isolate derived from Lp1A0 | This study |
| Lp1A6 | Single-spore isolate derived from Lp1A0 | This study |
| Lp1A3b | Single-spore isolate derived from Lp1A3 | This study |
| Lp1A3d | Single-spore isolate derived from Lp1A3 | This study |
| Lp1A6b | Single-spore isolate derived from Lp1A6 | This study |
| Lp1A6d | Single-spore isolate derived from Lp1A6 | This study |
| Lp1C3 | Single-spore isolate derived from Lp1C0 | This study |
| Lp1C6 | Single-spore isolate derived from Lp1C0 | This study |
| Lp1C3b | Single-spore isolate derived from Lp1C3 | This study |
| Lp1C3d | Single-spore isolate derived from Lp1C3 | This study |
| Lp1C6b | Single-spore isolate derived from Lp1C6 | This study |
| Lp1C6d | Single-spore isolate derived from Lp1C6 | This study |
| Lp1D3 | Single-spore isolate derived from Lp1D0 | This study |
| Lp1D6 | Single-spore isolate derived from Lp1D0 | This study |
| Lp1D3b | Single-spore isolate derived from Lp1D3 | This study |
| Lp1D3d | Single-spore isolate derived from Lp1D3 | This study |
| Lp1D6b | Single-spore isolate derived from Lp1D6 | This study |
| Lp1D6d | Single-spore isolate derived from Lp1D6 | This study |
| Lp1F3 | Single-spore isolate derived from Lp1F0 | This study |
| Lp1F6 | Single-spore isolate derived from Lp1F0 | This study |
| Lp2G3 | Single-spore isolate derived from Lp2G0 | This study |
| Lp2G6 | Single-spore isolate derived from Lp2G0 | This study |
| Lp2G3b | Single-spore isolate derived from Lp2G3 | This study |
| Lp2G3d | Single-spore isolate derived from Lp2G3 | This study |
| Lp2G6b | Single-spore isolate derived from Lp2G6 | This study |
| Lp2G6d | Single-spore isolate derived from Lp2G6 | This study |

Table 2.1 continued

| Strain, λ clone or plasmid | Relevant characteristics | Source or reference |
|---|---|---|
| *Neotyphodium lolii* (LpTG-1) | | |
| Lp5 | Putative Lp1 progenitor | M.J. Christensen |
| Lp9 | Atypical *N. lolii* LpTG-1 isolate | (Moon *et al.*, 1999) |
| *Epichloë typhina* (MP-1) | | |
| E8 | Putative Lp1 progenitor | M.J. Christensen |
| *Neotyphodium uncinatum* | | |
| Fp4 | Interspecific hybrid | C.D. Moon |
| *Neotyphodium coenophialum* | | |
| Tf27 | Interspecific hybrid | C.D. Moon |
| Tf28 | Interspecific hybrid | C.D. Moon |
| *Neotyphodium* sp. (FaTG-2) | | |
| Tf13 | Interspecific hybrid | C.D. Moon |
| Tf15 | Interspecific hybrid | C.D. Moon |
| *Neotyphodium* sp. FaTG-3 | | |
| Tf16 | Interspecific hybrid | C.D. Moon |
| *Epichloë festucae* | | |
| Frc7 | Putative sexual ancestor of LpTG-1 | C.D. Moon |

# Table 2.1 continued

| Strain, λ clone or plasmid | Relevant characteristics | Source or reference |
|---|---|---|
| **Bacterial cultures** | | |
| *Escherichia coli* | | |
| MC1022 | *ara*D Δ(*ara leu*) Δ(*lac*Z) M15 *gal*U *gal*K *str*A | |
| XL1-Blue | *sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A46 *thi rel*A1 *lac*⁻ F'[*pro*AB⁺ *lac*I$^q$ *lac*ZΔM15 Tnl0 (tet$^r$)] | (Bullock *et al.*, 1987) |
| PN1322 | MC1022 containing pPN49 | K. Marriott |
| PN1376 | XL1-Blue containing pPN50 | (Ganley, 1993) |
| **Phage λ Clones** | | |
| λPN1 | λEMBL3A clone containing an rDNA unit from Lp1 | C.A. Young |
| λMC11 | λEMBL3A clone containing *pyr*4-2 from Lp1 | (Collett *et al.*, 1995) |
| **Plasmids** | | |
| pUC118 | 3.2 kb Amp$^R$ | (Vieira and Messing, 1987) |
| pGEM®-T | 3.0 kb Amp$^R$ | Promega |
| pPN49 | pUC118 containing 5.6 kb *Sal*I *rrn* coding region λPN1 fragment | K. Marriott |
| pPN50 | pUC118 containing 4.1 kb *Sal*I IGS λPN1 fragment | (Gan, 1992) |

[a] For Lp1 and Lp2 the succeeding letter refers to the particular laboratory culture, with O representing the originally-isolated culture. The number succeeding this letter (3 or 6) refers to the first-round Single-spore isolate. The letter succeeding this number (b or d) refers to the second round Single-spore isolate.

### 2.2.1 POTATO DEXTROSE MEDIA (PD)

PD broth contains 2.4% (w/v) potato dextrose broth (Difco). The pH was adjusted to 6.5 prior to autoclaving. PD agar was prepared by the addition of agar (Davis) to a final concentration of 1.5% (w/v), unless otherwise indicated.

### 2.2.2 LURIA BROTH (LB) MEDIA

LB media contained 1% (w/v) tryptone (Difco), 0.5% (w/v) yeast extract (Difco) and 0.5% (w/v) NaCl. The pH was adjusted to 7.0 prior to autoclaving. LB agar was prepared by addition of agar (Davis) to a final concentration of 1.5% (w/v). LB media were supplemented with the antibiotic ampicillin to a final concentration of 100 µg/ml where appropriate.

### 2.2.3 SOC MEDIA

SOC media (Dower *et al.*, 1988) contained 20 mM glucose, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4.7H_2O$, 10 mM NaCl, 2% (w/v) tryptone (Difco) and 0.5% (w/v) yeast extract (Difco).

### 2.2.4 COMMON BUFFERS AND SOLUTIONS

**10 x Denhardt's**

10 x Denhardt's (Southern, 1975) contained (per litre): 50 ml of 1 M HEPES (BRL) at pH 7.0; 150 ml of 20 x SSC; 18 mg of phenol-extracted herring sperm DNA (Sigma); 1.0 g of SDS; 20 mg of *E. coli* tRNA (Sigma); 2 g of Ficoll (Sigma 70); 2 g of bovine serum albumin (Sigma); and 2 g of polyvinylpyrrolidone (Sigma PVP-10).

**Acrylamide mix**

Acrylamide mix for sequencing gels was prepared as follows: 288 g of urea, 34.2 g of acrylamide and 1.8 g of bis-acrylamide were made up to 500 ml and deionised with 5 g of Amberlite MB-3 (Sigma) for at least 30 min. This was filtered through a porous sintered glass funnel and made up to a final volume of 600 ml with 60 mls of 10 x sequencing TBE (this Section) and water.

**Alkaline lysis solutions**

| | |
|---|---|
| Solution I: | 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM $Na_2EDTA$ (pH 8.0). |
| Solution II: | 0.2 M NaOH, 1% (w/v) SDS. |
| Solution III: | 29.44 g potassium acetate and 11.5 ml glacial acetic acid in 100 ml. |

**Blotting solutions**

| | |
|---|---|
| Solution I: | 0.25 M HCl. |
| Solution II: | 0.5 M NaOH, 0.5 M NaCl. |
| Solution III: | 2.0 M NaCl, 0.5 M Tris-HCl, to pH 7.4. |
| 20 x SSC: | 5M NaCl, 0.3 M sodium citrate. |
| 2 x SSC: | 0.5 M NaCl, 0.03 M sodium citrate. |

**Dialysis tubing**

Dialysis tubing was prepared (Sambrook *et al.*, 1989) by boiling in 2% (w/v) sodium bicarbonate and 1 mM $Na_2EDTA$ (pH 8.0). It was rinsed well in MilliQ water, boiled for a further 10 min in 1 mM $Na_2EDTA$ (pH 8.0) and then stored at 4°C.

**DNA extraction buffer**

DNA extraction buffer contained 100 mM LiCl, 10 mM $Na_2EDTA$ (pH 8.0), 10 mM Tris-HCl (pH 8.0) and 0.5% (w/v) SDS.

**Exonuclease III digestion solutions**

| | |
|---|---|
| 1 x exoIII buffer: | 66 mM Tris-HCl (pH 8.0) and 0.66 mM $MgCl_2$ |
| 7.4 x S1 nuclease buffer: | 0.3 M potassium acetate (pH 4.6), 2.5 M NaCl, 10 mM $ZnSO_4$ and 50% (v/v) glycerol |
| S1 nuclease mix: | 2.6 µl of 7.4 x S1 nuclease buffer and 4 U of S1 nuclease (Boehringer Mannheim) per 19 µl |
| S1 stop buffer: | 0.3 M Tris-HCl (pH 8.0) and 50 mM $Na_2EDTA$ (pH 8.0) |
| ExoIII ligation mix: | 2.3 µl of 10 x T4 DNA ligase ligation buffer, 4 U T4 DNA ligase (both New England Biolabs) and 2.3 µl of 50% (w/v) PEG-6,000 per 38 µl |

**Junction fragment cloning solutions**

| | |
|---|---|
| 10 x annealing buffer: | 100 mM Tris-HCl (pH 7.6), 1 M NaCl and 10 mM $Na_2EDTA$ (pH 8.0) |
| 100 mM sodium phosphate (pH 7.4): | 9.5 ml of 100 mM $NaH_2PO_4$ and 40.5 ml of 100 mM $Na_2HPO_4$ in 100 ml |
| PBS/BSA buffer: | 10 mM sodium phosphate (pH 7.4), 150 mM NaCl and 0.1% (w/v) bovine serum albumin (Sigma) |
| 2 x DBW buffer: | 10 mM Tris-HCl (pH 8.0), 1 mM $Na_2EDTA$ (pH 8.0) and 2 M NaCl |

**Lysozyme**

Lysozyme (Sigma) was made up to 10 mg/ml in 10 mM Tris-HCl (pH 8.0) and stored at -20°C.

**Protoplasting and CHEF agarose plug preparation solutions**

All protoplasting and CHEF agarose plug preparation solutions were autoclaved prior to use.

| | |
|---|---|
| OM buffer: | 1.5 M $MgSO_4.7H_2O$, 10 mM $Na_2HPO_4$ and the pH was adjusted to 5.8 with 100 mM $NaH_2PO_4.2H_2O$. |
| ST buffer: | 0.6 M sorbitol and 100 mM Tris-HCl (pH 8.0). |
| STC buffer: | 1 M sorbitol, 50 mM $CaCl_2$ and 50 mM Tris-HCl (pH 8.0). |
| GMB buffer: | 0.9 M sorbitol and 0.125 M $Na_2EDTA$ (pH 8.0). |
| LMP in GMB: | 1.6% (w/v) low melting point (LMP) agarose (Sigma) in GMB buffer. |
| SE buffer: | 0.25 M $Na_2EDTA$ (pH 8.0), and 2% (w/v) SDS after autoclaving. |
| 10 x ET buffer: | 10 mM Tris-HCl (pH 8.0), 0.5 M $Na_2EDTA$ (pH 8.0), then 1% (w/v) SLS and 20 mg Proteinase K (Boehringer Mannheim) after autoclaving. |
| 1 x ET buffer: | 1 mM Tris-HCl (pH 8.0) and 50 mM $Na_2EDTA$ (pH 8.0). |

**RNase (DNase free)**

RNaseA (Sigma) prepared at 10 mg/ml in 10 mM Tris-HCl (pH 7.5) and 15 mM NaCl was heated to 100°C for 15 min, and allowed to cool slowly to room temperature. It was stored at -20°C and was used at 0.2 mg/ml to treat DNA.

**SDS loading dye**

SDS loading dye contained 1% (w/v) SDS, 0.02% (w/v) bromophenol blue, 20% (w/v) sucrose and 5 mM $Na_2EDTA$ (pH 8.0). 50% (v/v) SDS loading dye was mixed with DNA solutions to be separated on mini gels (Section 2.8.2), and 25% (v/v) SDS loading dye was mixed with DNA solutions to be separated on overnight gels (Section 2.8.3).

**SSC-saturated isopropanol**

Equal volumes of 20 x SSC (this Section) and isopropanol were stirred overnight. The top layer was used for extracting ethidium bromide from DNA solutions.

**STET buffer**

STET buffer contained 8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM Tris-HCl (pH 8.0) and 50 mM $Na_2EDTA$ (pH 8.0).

**TAE buffer**

At 1 x concentration TAE buffer contained 40 mM Tris-acetate and 2 mM $Na_2EDTA$ at pH 8.5.

**TBE buffer**

At 1 x concentration TBE buffer contained 89 mM Tris, 89 mM boric acid and 2.5 mM $Na_2EDTA$ at pH 8.2.

**TBE sequencing buffer**

At 1 x concentration TBE sequencing buffer contained 134 mM Tris, 45 mM boric acid and 2.5 mM $Na_2EDTA$ at pH 8.8.

**Tris/EDTA buffers (TE and TES)**

| | |
|---|---|
| TE (10/1): | 10 mM Tris-HCl (pH 8.0) and 1 mM $Na_2EDTA$ (pH 8.0). |
| TE (10/0.1): | 10 mM Tris-HCl (pH 8.0) and 0.1 mM $Na_2EDTA$ (pH 8.0). |
| TES (10/1/100) | 10 mM Tris-HCl (pH 8.0), 1 mM $Na_2EDTA$ (pH 8.0) and 100 mM NaCl. |

**Tris-equilibrated phenol**

Tris-equilibrated phenol was supplied by USB. 8-hydroxyquinoline was added to give a final concentration of 0.1% (w/v), and the equilibrated phenol was stored at 4°C.

## 2.3 DNA ISOLATION

### 2.3.1 ISOLATION OF FUNGAL GENOMIC DNA

DNA was extracted from fungal cultures using the method of Brownlee (1988). Mycelia from PD broth cultures (Section 2.1.2) were harvested by filtration through 11 cm Whatman 1 filter paper under vacuum using a funnel and Buchner flask, frozen in liquid $N_2$ and then freeze-dried. The lyophilised mycelium was ground to a fine powder in a mortar and pestle under liquid $N_2$ and resuspended thoroughly in 8 ml of DNA extraction buffer (Section 2.2.4). 6 ml of Tris-equilibrated phenol (Section 2.2.4) was added, and this was centrifuged at 20,000 *g* for 15 min at 4°C. The aqueous phase was extracted with 6 ml of Tris-equilibrated phenol and 6 ml of chloroform, and then with 6 ml of chloroform, with centrifugation performed as above. The aqueous phase was centrifuged at 48,000 *g* for 25 mins at 4°C, the supernatant removed and the DNA was precipitated by addition of 2 volumes of 95% ethanol. The DNA was pelleted by centrifugation at 20,000 *g* for 20 min at 4°C. The pellet was washed in 70% ethanol, dried and resuspended in 500 µl of TE (10/0.1) buffer (Section 2.2.4).

### 2.3.2 EXTRACTION OF PLASMID DNA FROM BACTERIAL CULTURES BY THE RAPID BOIL METHOD

Plasmid DNA was isolated from *E. coli* by the method of Holmes and Quigley (1981). The cells from 1.5 ml of a 5 ml LB broth culture (Section 2.1.4) were pelleted at 13,000 *g* in a microcentrifuge and resuspended in 350 µl of STET buffer (Section 2.2.4). 25 µl of lysozyme (Section 2.2.4) was added, the solution boiled for 40 sec and centrifuged at 13,000 *g* for 10 min. The gelatinous pellet was removed and the DNA precipitated by addition of 2.5 volumes of 95% ethanol or 0.6 volumes of isopropanol at -20°C for at least 2 h. The DNA was pelleted at 13,000 *g* in a microcentrifuge, washed with 70% ethanol, dried and resuspended in 50 µl of MilliQ $H_2O$.

### 2.3.3 EXTRACTION OF PLASMID DNA FROM BACTERIAL CULTURES BY THE ALKALINE LYSIS METHOD

Plasmid DNA from *E. coli* to be sequenced was prepared by the method of Sambrook *et al.* (1989). Cells from 3 ml of a 5 ml LB broth culture (Section 2.1.3) were pelleted at 13,000 *g* in a microcentrifuge and resuspended in 100 µl of solution I (Section 2.2.4). 200 µl of solution II (Section 2.2.4) was added and the contents of the tube mixed thoroughly by inversion. Then 150 µl of solution III (Section 2.2.4) was added, and the mixture vortexed in an inverted position. This was incubated on ice for 5 min, and centrifuged at 13,000 *g* for 5 min at 4°C. The supernatant was extracted with an equal volume of Tris-equilibrated phenol (Section 2.2.4) and this was centrifuged at 13,000 *g* for 5 min at 4°C. The aqueous phase was extracted with equal volumes of Tris-equilibrated phenol/chloroform, then an equal volume of chloroform, with centrifugation performed as before. The DNA was precipitated by addition of 2.5 volumes of 95% ethanol or 0.6 volumes of isopropanol at -20°C for two hours. The DNA was pelleted at 13,000 *g* in a microcentrifuge, washed with 70% ethanol, dried and resuspended in 50 µl of MilliQ water.

### 2.3.4 EXTRACTION OF PLASMID DNA FROM BACTERIAL CULTURES BY THE CAESIUM CHLORIDE METHOD

High purity plasmid DNA from *E. coli* was prepared by an alkaline lysis/CsCl method modified from Ish-Horowitz and Burke (1981). A 5 ml LB broth culture (Section 2.1.4) was used to inoculate 500 ml of LB broth (Section 2.2.2), and this was grown for 16 h at 37°C shaking at 200 rpm. The cells were harvested by centrifugation at 10,000 *g* for 10 min at 4°C, washed with a total of 100 ml of TE (10/1) buffer (Section 2.2.4), and re-harvested as before. The cells were resuspended in 30 ml of solution I (Section 2.2.4) containing 150 mg of lysozyme (Boehringer Mannheim) and incubated at room temperature for 10 min. 60 ml of solution II (Section 2.2.4) was added, mixed by inversion and the mixture incubated on ice for 10 min. 45 ml of solution III (Section 2.2.4) was added, mixed by inversion and the mixture incubated for a further 10 min on ice. This was centrifuged at 10,000 *g* for 10 min at 4°C. Isopropanol (0.6 volumes) was added to the supernatant and the mixture was incubated on ice for 20 min. The DNA was pelleted by centrifugation at 23,500 *g* for 10 min at 4°C, washed with 70% ethanol, dried and resuspended in 7.5 ml of TE (10/1) buffer. CsCl (Boehringer Mannheim) was added to the DNA solution in a ratio of 1.05 g of CsCl/ml of DNA solution, followed by ethidium bromide in a ratio of 75 µl of 10 mg/ml ethidium bromide per ml of DNA/CsCl solution. The solution was mixed well, left overnight at 4°C, and then centrifuged at 17,000 *g* for 10 min at 4°C. The supernatant was checked for a refractive index of 1.3920 and adjusted with either TE (10/1) buffer or CsCl if necessary. The DNA was ultracentrifuged at 166,000 *g* for 15 h at 18°C and the resulting band corresponding to the plasmid DNA was removed with an 18-gauge hypodermic needle and syringe. The ethidium bromide was removed by repeated extraction with an equal volume of SSC-saturated isopropanol (Section 2.2.4) until the DNA solution was clear. The CsCl was

removed by dialysis (dialysis tubing; Section 2.2.4) against TES (10/1/100) buffer (Section 2.2.4). The resulting DNA was quantified by spectrophotometric analysis (Section 2.4.3) and stored at 4°C in TES (10/1/100).

## 2.4 DNA QUANTIFICATION

### 2.4.1 DNA QUANTIFICATION BY ETHIDIUM BROMIDE STAINING

A sample of DNA was loaded on a mini gel alongside a series of standards of known DNA concentration and these were separated by electrophoresis (Section 2.8.2). The mini gel was stained in ethidium bromide (Section 2.8.4), and the intensity of fluorescence of the unknown sample and concentration standards were compared. The three different types of concentration standards used are described below.

| Type of standard | Range of standard | Sample DNA size used for |
|---|---|---|
| Uncut phage λ | 10 ng - 200 ng | Genomic DNA |
| Linearised pUC118 | 2.5 ng - 20 ng | Fragments in the 1.5 kb - 10 kb size range |
| Low mass ladder (Gibco BRL) | 10 ng - 100 ng | Fragments less than 2 kb |

### 2.4.2 FLUOROMETRIC QUANTITATION

DNA was quantitated in a dye solution containing 1 x TNE buffer (10 mM Tris-HCl (pH 7.4), 1 mM $Na_2EDTA$ and 100 mM NaCl) and 0.1 μg/ml of Hoechst 33258 dye. A Hoefer Scientific TKO 100 fluorometer was used. The scale of the fluorometer was zeroed with the dye solution alone until a steady state was achieved, and was then set to 100 using 2 μl of 100 μg/ml calf thymus DNA added to 2 ml of the dye solution. Once the scale was reliably set, 2μl of sample DNA was added to 2 ml of the dye solution, with the reading from the fluorometer giving the concentration of the sample DNA in ng/μl.

### 2.4.3 SPECTROPHOTOMETRIC QUANTITATION

Spectrophotometric quantitation was used for DNA isolated by CsCl extraction (Section 2.3.4). Absorbance readings from the spectrophotometer were taken from the DNA sample in quartz cuvettes at 260 nm and 280 nm. The reading at 260 nm allowed calculation of the concentration of nucleic acid present in the sample - an absorbance of 1 corresponds to approximately 50 μg/ml of double-stranded DNA. The absorbance ratio, $A_{260}/A_{280}$, was used as an estimate of DNA purity, as pure DNA has a ratio of 1.8.

## 2.5 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction endonuclease (RE) digests were carried out as per the manufacturer's recommendation using a 2 - 5 fold excess of RE endonuclease. Genomic digests were supplemented with 100 µg/ml of acetylated bovine serum albumin (Promega). In general genomic DNA was digested for 16 h, and plasmid and PCR product DNA was digested for 2 h. A small aliquot of the digested DNA was checked on a mini gel (Section 2.8.2) to ensure complete digestion. If digestion was not complete, more RE endonuclease was added, and the reaction incubated for another 2 h (genomic DNA) or 1 h (plasmid or PCR product DNA). DNA recalcitrant to digestion was supplemented with 100 pmol of spermidine (Sigma) to aid digestion (Pingoud *et al.*, 1984). If digestion was still incomplete the DNA was purified by phenol/chloroform extraction (Section 2.6) and ethanol precipitation (Section 2.7). Once complete the digestion was stopped by the addition of SDS loading dye (Section 2.2.4).

## 2.6 PHENOL/CHLOROFORM EXTRACTION

DNA was purified by phenol/chloroform extraction to inactivate and remove enzymes and other impurities that may interfere with further manipulations. The DNA solution was extracted with an equal volume of Tris-equilibrated phenol (Section 2.2.4) and chloroform by vortexing, and centrifuged at 13,000 *g* for 5 min in a microcentrifuge. The aqueous phase was then extracted with 1 volume of chloroform as before. Finally the DNA was ethanol/isopropanol precipitated as described in Section 2.7.

## 2.7 DNA PRECIPITATION

DNA to be precipitated was mixed with 0.1 volumes of 3 M sodium acetate (pH 7.0) and either 2.5 volumes of 95% ethanol or 0.6 volumes of isopropanol. This mixture was left at -20°C for 2 h. The DNA was then pelleted by centrifugation at 13,000 *g* for 15 min in a microcentrifuge. The DNA pellet was rinsed with 70% ethanol, dried and resuspended in either $H_2O$ or TE (10/0.1) buffer (Section 2.2.4) as appropriate. Where possible ethanol was used for DNA precipitation as solutes (such as NaCl) are less easily coprecipitated.

# 2.8 AGAROSE GEL ELECTROPHORESIS

## 2.8.1 AGAROSE GELS

Agarose was made up in 1 x TBE buffer (Section 2.2.4) at the concentrations best suited for the desired separation as described below, unless indicated otherwise. The agarose was melted in a microwave and allowed to equilibrate to 50°C in a water bath before pouring.

| Grade of agarose | % (w/v) of agarose | Size range of separation |
|---|---|---|
| Molecular Biology (Gibco BRL) | 0.7 | 2.0 kb - 25 kb |
| Molecular Biology (Gibco BRL) | 1.5 | 0.5 kb - 3.0 kb |
| NuSieve (FMC) | 2.0 | 0.3 kb - 1.5 kb |
| NuSieve (FMC) | 3.0 | 0.1 kb - 1.0 kb |

## 2.8.2 MINI GEL ELECTROPHORESIS

Horizontal agarose mini gels were run on home-made mini gels or Horizon® 58 mini gels (Bio-Rad) at 60 - 100 volts in 1 x TBE buffer (Section 2.2.4) unless indicated otherwise.

## 2.8.3 OVERNIGHT GELS

Overnight gels were run on a Bio-Rad DNA Sub Cell™ (150 x 200 mm) gel box at 35 volts.

## 2.8.4 DETECTION OF DNA

DNA was detected after electrophoresis by staining the gel in ethidium bromide (1 µg/ml ethidium bromide in reverse osmosis $H_2O$) for 10 min for a mini gel or 30 min for an overnight, Seaplaque (Section 2.9) or CHEF gel (Section 2.19.4), followed by destaining in water. The bands were visualised on a UV transilluminator and photographed with Polaroid 667 film or an Alpha Innotech gel documentation system.

## 2.8.5 DETERMINATION OF DNA BAND SIZES

The sizes of DNA bands were determined after agarose gel electrophoresis by measuring the distance (in mm) a fragment had migrated from the well. The size was calculated by interpolation from a plot of known bands from one of the size standards described below using either the Gel_Frag_Sizer Macintosh programme (dog Star Software) or Alpha Innotech gel documentation software.

| Type of size standard | Range of sizes |
|---|---|
| λ *Hin*dIII ladder | 2.0 kb - 23 kb |
| 1 kb ladder (Gibco BRL) | 1.0 kb - 10 kb |
| 100 bp ladder (Gibco BRL) | 100 bp - 1.0 kb |

## 2.9 SEAPLAQUE DNA EXTRACTIONS

To extract individual fragments from an agarose gel 0.7% (w/v) Seaplaque agarose (FMC) made up with 1 x TAE buffer (Section 2.2.4) was used. Electrophoresis was performed using a Bio-Rad Wide Mini-Sub™ Cell gel box in 1 x TAE buffer at 60 volts. After separation the gel was stained in ethidium bromide (Section 2.8.4) and the band of interest was excised under long wavelength UV light. This agarose was melted at 65°C for 10 min, mixed with an equal volume of Tris-equilibrated phenol (Section 2.2.4), vortexed and frozen at -20°C for at least 2 h. The sample was thawed at room temperature and centrifuged at 13,000 *g* for 10 min in a microcentrifuge. The aqueous phase was then phenol/chloroform extracted (Section 2.6) and the DNA precipitated (Section 2.7). Finally the concentration of the excised fragment was determined (Section 2.4).

## 2.10 SOUTHERN BLOTTING AND HYBRIDISATION

### 2.10.1 SOUTHERN (CAPILLARY) BLOTTING

The blotting method used was based on that of Southern (1975). DNA to be transferred onto a nylon membrane was separated by overnight (Section 2.8.3) or CHEF (Section 2.19.4) gel electrophoresis and photographed as described in Section 2.8.4. The gel was gently agitated in the following series of solutions (Section 2.2.4): blotting solution I for 15 min; blotting solution II for 30 min, and blotting solution III for 1 h. The gel was rinsed for 2 min in 2 x SSC (Section 2.2.4) and assembled onto a blotting stand as follows. Two sheets of 3MM paper (Whatman) soaking in 20 x SSC (Section 2.2.4) were used as wicks, and the sheets were covered with glad wrap with a hole cut slightly smaller than the gel area. The treated gel was then placed on the stand over this hole, and a piece of nylon membrane (Hybond N from Amersham or Nylon membrane from Boehringer Mannheim), cut slightly larger than the gel and wetted in 2 x SSC, was placed on top. Two sheets of 3MM paper, slightly smaller than the gel and soaked in 2 x SSC, were placed on top of this, followed by two dry pieces and finally a stack of paper towels with a weight on top. After overnight transfer of the DNA from the gel to the nylon membrane, the membrane was removed, washed for 5 min in 2 x SSC and baked for 2 h at 80°C under vacuum to fix the DNA to the membrane.

### 2.10.2 MAKING RADIOLABELLED DNA PROBES

Probes were radiolabelled to a high specific activity with [$\alpha$-$^{32}$P]dCTP (3,000 Ci/mmole, Amersham) using a High Prime (Boehringer Mannheim) random priming kit according to manufacturer's instructions. Unincorporated nucleotides were removed by centrifuging the probe through a Sephadex G-50 column (ProbeQuant™ G-50 Micro Column, Pharmacia Biotech) at 735 *g* for 2 min. Radiolabelled probes were denatured by boiling and rapid cooling prior to hybridisation.

### 2.10.3 HYBRIDISATION OF RADIOLABELLED PROBES

DNA hybridisations were performed in roller bottles in a Bachofer hybridisation oven. Nylon membranes (Section 2.10.1) were pre-hybridised with ~ 20 ml of 10 x Denhardt's solution (Section 2.2.4) for 2 h at the hybridisation temperature prior to addition of the probe. The 10 x Denhardt's solution was removed to leave ~ 5 ml and then the probe(s) (Section 2.10.2) was added. Hybridisation was performed overnight at 65°C unless indicated otherwise. Following hybridisation three sets of washes were performed, each at room temperature in 2 X SSC (Section 2.2.4) for 15 minutes.

### 2.10.4 DETECTION OF HYBRIDISATION SIGNAL

After hybridisation and washing the nylon membrane was sealed in glad wrap and placed in a light-sealed autoradiographic cassette with a piece of X-ray film (Fuji New RX 35 x 43 cm or 18 x 24 cm; or Kodax X-OMAT™ AR 35 x 43 cm). Exposure was performed at -70°C with an intensifying screen. The X-ray film was then developed using standard procedures.

### 2.10.5 STRIPPING NYLON MEMBRANES

Radiolabelled probe was stripped from the nylon membrane by pouring 400 ml of boiling 0.1% (w/v) SDS onto the membrane and agitating this until the solution had cooled to room temperature. This was repeated as often as necessary. The nylon membrane was checked either by overnight exposure to X-ray film (Fuji New RX 35 x 43 cm or 18 x 24 cm; or Kodax X-OMAT™ AR 35 x 43 cm) or by a Series 900 Mini-monitor Geiger counter (Mini-Instruments).

## 2.11 CAP TREATMENT OF VECTORS

Restriction endonuclease digested pUC118 DNA (Section 2.5) was incubated with 0.1 unit of calf intestine alkaline phosphatase (CAP; Boehringer Mannheim) at 37°C for 30 min. $Na_2EDTA$ (pH 8.0; final concentration of 5 mM), SDS (final concentration of 0.5% [w/v]) and proteinase K (final concentration of 50 µg/ml; Boehringer Mannheim) were added and the mixture was

incubated at 56°C for 30 min to inactivate the CAP. The DNA was purified by phenol/chloroform extraction (Section 2.6), precipitated (Section 2.7) and then diluted to 10 - 20 ng/µl with MilliQ $H_2O$.

## 2.12 LIGATIONS

### 2.12.1 LIGATION INTO pUC118

Ligations were performed in 20 µl reaction volumes unless otherwise indicated. The reactions contained the following: 2 µl of ligation buffer (New England Biolabs), 20 ng of vector (Section 2.11), 0.1 µl of T4 DNA ligase (40 U; New England Biolabs) and a 2 - 3 molar excess of insert:vector unless otherwise indicated. The ligation mixture was incubated overnight at 4°C. To check for successful ligation, 2 µl samples of the ligation reaction were removed before and after addition of the T4 DNA ligase, and these were compared by separation on an agarose mini gel (Section 2.8.2). Following successful ligation, the ligation reactions were transformed into *E. coli* (Section 2.13.2).

### 2.12.2 LIGATION INTO pGEM®-T

PCR products (Section 2.17) to be cloned were ligated into the pGEM®-T T-tailed vector (Promega). Ligations were performed in 10 µl reaction volumes according to manufacturer's instructions, except that only half the recommended amount of vector and insert DNA were added. The ligation mixture was incubated overnight at 4°C. Ligations were not checked on a gel before transformation into *E. coli* (Section 2.13.2).

## 2.13 TRANSFORMATION OF *E. COLI*

*E. coli* XL1-Blue cells (Bullock *et al.*, 1987) were transformed by the electroporation method.

### 2.13.1 PREPARATION OF ELECTROCOMPETENT CELLS

One litre of LB broth (Section 2.2.2) was inoculated with XL1-Blue cells and grown at 37°C with vigorous shaking to mid-log phase ($A_{600}$ 0.5 - 1.0, approximately 3 h). The cells were chilled on ice for 20 min and harvested by centrifugation at 4,000 *g* for 10 min at 4°C. The cells were resuspended twice in ice-cold water (1 litre then 500 ml, centrifuging as before) followed by resuspension in ice-cold 10% (v/v) glycerol (20 ml then 4 ml, centrifuging as before). The cells were then stored at -70°C in 200 µl aliquots.

### 2.13.2 ELECTROPORATION AND DETECTION OF TRANSFORMANTS

Electrocompetent *E. coli* XL1-Blue cells (Section 2.13.1) were thawed at room temperature, and 40 µl were mixed with 2 µl of a ligation reaction (Section 2.12) and incubated on ice for 1 min. The cells were placed in the base of an ice-cold 0.2 cm electroporation cuvette (Bio-Rad). This was pulsed with the following settings in a Bio-Rad Gene Pulser™: 25 µF, 2.5 kilovolts and 200 ohms. When pulsing was complete 1 ml of SOC medium (Section 2.2.3) was immediately added to the cells, and these were incubated at 37°C for 1 h. Positive (2 ng of uncut pUC118 (Table 2.1) DNA) and negative (no DNA) controls were always employed. The cells were diluted as necessary in SOC media, plated onto LB agar plates supplemented with ampicillin (Section 2.1.4) and grown for 16 h at 37°C. Any colonies that grew were tested for the correct insert either by plasmid isolation (Section 2.3.2 or Section 2.3.3) and restriction endonuclease digestion (Section 2.5) or by using a small number of cells in a PCR reaction (Section 2.17.2) in the place of DNA with appropriate primers.

## 2.14 SUBCLONING DNA FRAGMENTS

Subcloning of a PCR product (Section 2.17) digested with a restriction endonuclease (Section 2.5) at one end before cloning (Section 3.1.10) was performed as follows. The ends were filled in by incubation with 5 mM dithiothreitol, 100 µM dNTPs and 1 U of T4 DNA polymerase (all Boehringer Mannheim) at 37°C for 30 min. This was phenol/chloroform extracted (Section 2.6) and ethanol precipitated (Section 2.7). The PCR product was then A-tailed by incubation in a 20 µl solution containing 1 x *Taq* DNA polymerase buffer, 5 U of *Taq* DNA polymerase, 0.2 mM dATP (all Boehringer Mannheim) and 2.5 mM $MgCl_2$ for 30 min at 70°C. These A-tailed PCR products were then ligated into the pGEM-T vector (Section 2.12.2).

## 2.15 GENERATION OF NESTED DELETION BY EXONUCLEASE III DIGESTION

Exonuclease III was used to create a set of nested deletions in the Lp1 rDNA 4.1 kb *Sal*I IGS insert of pPN50 to facilitate sequencing of this clone. The generation of nested deletions by exonuclease III was performed using a modification of the methods from Ausubel *et al.* (1987-1993) and the Erase-A-Base® system (Promega) as follows.

5 µg of CsCl-prepared (Section 2.3.4) pPN50 DNA was digested with the restriction endonucleases *Bam*HI and *Sst*I (Section 2.5) to generate a 5' overhang adjacent to the 4.1 kb *Sal*I IGS insert and a 3' overhang on the opposite side of the double-strand break adjacent to the M13-reverse sequencing primer site, respectively. This ensures exonuclease III digestion occurs unidirectionally into the insert, as 3' overhangs are resistant to exonuclease III digestion. The

DNA was phenol/chloroform extracted (Section 2.6), ethanol precipitated (Section 2.7) and resuspended in 25 µl of 1 x exoIII buffer (Section 2.2.4) to give a final concentration of 0.2 µg/µl. The amount of exonuclease III used is determined by the concentration of susceptible ends, with 150 U of exonuclease III used per pmol of susceptible ends. On the basis of the relation: 1 µg of 1 kb DNA = 1.5 pmol of susceptible ends, there were a total of 1.1 pmol of susceptible ends in 5 µg of pPN50 DNA. Therefore 175 U of exonuclease III (Boehringer Mannheim) was added to the DNA which had been preincubated at 37°C. Incubation was performed at 37°C, with 3 µl aliquots removed every 70 sec up to 630 sec (for a total of nine time points) and stored on ice. To each of these samples 19 µl of S1 nuclease mix (Section 2.2.4) was added immediately, and these were incubated at room temperature for 30 min to remove any single stranded tails. 5 µl of S1 stop buffer (Section 2.2.4) was added and incubated at 70°C for 10 mins. At this point 7 µl of each time point were loaded on an agarose mini gel (Section 2.8.2) to check the digestion (see Figure 3.7A). To ensure blunt ends, 2 U of Klenow DNA polymerase (Boehringer Mannheim) and 1 µl of each of the four dNTPs (Boehringer Mannheim) at 0.25 mM were added to the reaction mixtures from each time point, and these were incubated at 37°C for 10 min. Finally 38 µl of exoIII ligation buffer (Section 2.2.4) was added to each sample and incubated at 4°C for 16 hours.

Selection of clones involved electroporating (Section 2.13.2) each of the samples into electrocompetent XL1-Blue cells (Section 2.13.1). Plasmid DNA isolated from transformants by the rapid boil method (Section 2.3.2) was linearised with the restriction endonuclease *Hin*dIII (Section 2.5) and run on an agarose mini gel (Section 2.8.2). Clones were identified (Section 2.8.5) that had successive deletions differing by ~ 300 bp covering the entire 4.1 kb *Sal*I IGS insert. Plasmid DNA from the transformants harbouring these deletion clones was isolated using the alkaline lysis method (Section 2.3.3), and these were sequenced (Section 2.16.1) using the M13-reverse primer (Table 2.2).

## 2.16 DNA SEQUENCING

Sequencing of the nested deletion clones from Section 2.15 was performed using the Sequenase™ kit, whilst all subsequent sequencing was performed using Amplicycle™ sequencing or automatic sequencing.

### 2.16.1 SEQUENASE™

The Sequenase™ kit (USB) was used for sequencing alkaline lysis DNA preparations (Section 2.3.3). 5 µg of plasmid DNA was denatured with a freshly prepared solution of 2 M NaOH and 2 mM $Na_2EDTA$ in a 20 µl volume and incubated at room temperature for 5 min. The DNA was placed on ice and 8 µl of 1 M Tris-HCl (pH 9.0) and 3 µl of 3 M sodium acetate (pH 4.5) were added. The solution was mixed well and 75 µl of ice-cold 95% ethanol added. The DNA was

pelleted by centrifugation at 13,000 *g* for 10 min in a microcentrifuge, washed with 200 µl of 70% ethanol and dried. The DNA was resuspended in 7 µl of MilliQ $H_2O$ in a 0.5 ml tube, 1 µl of the M13-reverse sequencing primer (0.5 pmol; Table 2.2) and 2 µl of 5 x Sequenase reaction buffer added, and the mixture was incubated to allow annealing at 37°C for 30 min. After incubation the following reagents were added: 1 µl of 0.1 M dithiothreitol, 2 µl of Sequenase labelling mix (diluted in 5 x MilliQ $H_2O$), 0.5 µl of [$\alpha$-$^{35}$S]dATP (>1,000 Ci/mmol; Amersham) and 2 µl of Sequenase polymerase (diluted in 8 x Sequenase enzyme dilution buffer). This was incubated for 5 min at room temperature. 2.5 µl of the four Sequenase termination mixtures were pipetted into separate wells of a microtitre plate and 3.5 µl of the above labelling mixture was added to each. These were incubated for a further 5 min at 37°C. The reactions were stopped by addition of 4 µl of Sequenase stop solution and stored at -20°C.

### 2.16.2 AMPLICYCLE™ SEQUENCING

Sequencing using the Amplicycle™ sequencing kit (Perkin-Elmer) was performed on plasmid DNA prepared with the alkaline lysis method (Section 2.3.3) or purified PCR products (Section 2.17.4). A 30 µl reaction mixture containing 200 ng of plasmid DNA or 5 ng of PCR product, 4 µl of Amplicycle cycling mix, 1 µl of the appropriate sequencing primer (1 pmol) and 1 µl of [$\alpha$-$^{33}$P]dCTP (~ 2,500 Ci/mmol; Amersham) was prepared. 6 µl of this was mixed with 2 µl of each Amplicycle termination mix in 0.2 ml tubes and placed in a Corbett thermocycler set at 94°C. The following thermocycling programme was used: 94°C for 2 min, followed by 25 cycles of 94°C for 30 sec, then 50°C for 30 sec and 72°C for 1 min. After thermocycling was complete, 4 µl of Amplicycle stop solution was added to each sample and they were stored at 4°C.

### 2.16.3 ELECTROPHORESIS OF SEQUENCING REACTIONS

Radiolabelled sequencing reactions from 2.16.1 and 2.16.2 were separated by polyacrylamide gel electrophoresis. One glass plate used to form the sequencing gel was siliconised with Sigmacote (Sigma) prior to pouring. The gels were poured with 60 ml of acrylamide mix (Section 2.2.4) containing 36 µl of TEMED (USB) and 360 µl of 10% (w/v) ammonium persulphate, and left to set. Once set, gels were pre-run at 1,500 volts for 15 min in 1 x sequencing TBE (Section 2.2.4) on an S2 vertical sequencing apparatus (BRL). Radiolabelled sequencing reactions were denatured at 80°C for 2 min immediately prior to loading and 3 µl was loaded onto the gel. The samples were run for either a short run of ~2 h (until the first blue dye front ran off the bottom of the gel) or a long run of ~6 h (when the three dye fronts had run off) at a constant 1,500 volts. The gel was disassembled, fixed in 10% (v/v) acetic acid and 10% (v/v) 95% ethanol for 15 min, and dried for 2 h at 80°C under vacuum in a BioRad Model 583 gel dryer. Finally the gel was placed in a light-sealed autoradiographic cassette with a piece of X-ray film and exposed as set out in Section 2.10.4. The DNA sequence was read manually from the developed X-ray film.

### 2.16.4 AUTOMATIC SEQUENCING

Automatic sequencing using an ABI 377 automatic sequencer was performed on plasmid DNA prepared by the alkaline lysis method (Section 2.3.3) or purified PCR products (Section 2.17.4) using Prism (ABI) dye terminator chemistry according to manufacturer's instructions.

### 2.16.5 SEQUENCE ANALYSIS

The DNA sequence was analysed with the Wisconsin Package version 9.1, Genetics Computer Group (GCG), Madison, WI. The programmes used were the Fragment assembly package, GAP, MAP and BESTFIT. Homology searches to the Genbank DNA database were performed using BLAST (Altschul *et al.*, 1997) through Netscape Navigator at NCBI (www.ncbi.nlm.nih.gov/). Alignments were performed using the Clustal W(1.5) programme (Thompson *et al.*, 1994) or the GCG GAP programme.

## 2.17 POLYMERASE CHAIN REACTION (PCR)

### 2.17.1 OLIGONUCLEOTIDE PRIMERS

Oligonucleotide primers were synthesised by Gibco BRL. Each primer was rehydrated to a concentration of 100 µM and stored at -20°C. Primers were used at a stock concentration of 10 µM for PCR analysis (Section 2.17.2) and 1 µM for sequencing reactions (Section 2.16). The sequences of primers used in this study are presented in Table 2.2.

### 2.17.2 PCR CONDITIONS

PCR reactions were carried out in a final volume of 12.5 µl, 25 µl or 50 µl, all in 0.2 ml tubes. The reactions contained 1 x *Taq* DNA polymerase buffer (Boehringer Mannheim), 50 µM of each dNTP, 200 nM of each primer, 0.01 U/µl of *Taq* DNA polymerase (Boehringer Mannheim) and 0.2 ng/µl of genomic DNA or 40 pg/µl of plasmid DNA (final concentration), unless otherwise indicated. Expand™ High Fidelity DNA polymerase (Boehringer Mannheim) was used with templates recalcitrant to amplification, and these were also supplemented with either 2% or 5% (v/v) dimethyl sulphoxide (BDH; Shen and Hohn 1992). All thermocycling was performed in a Corbett thermocycler using the programmes described in Table 2.3 unless otherwise indicated. The reactions were separated on agarose mini gels (Section 2.8.2).

Table 2.2 Oligonucleotide Primers

| Primer name | Primer sequence |
|---|---|
| nts1 | CGGCTCTTCCTATCATACCGAAG |
| nts2 | GACTCCCCTCGGGATTAGCATAG |
| nts3 | TCTTGCAGACGTCTACTCCGTG |
| nts4 | GAGACAAGCATATGACTAC |
| nts7 | TGCGGGTGCGCTATCGAGATG |
| nts8 | GCAAATCACAGTCACCAGCGG |
| nts11 | ACACAACCTGGATCGGG |
| nts12 | GTGTCGCCGCATAGCCC |
| nts13 | CCGCTGGCCGTCATGCC |
| nts14 | TCGGGGAAGTCGACGCC |
| M13-forward | GTTTTCCCAGTCACGAC |
| M13-reverse | CAGGAAACAGCTATGAC |
| TAG | TTTGTCCGCTCGGTTGC |
| 111-L | TTTGTCCGCTCGGTTGCCGCGGGCAGAGTGGTGCC |
| 111-R | TTTGTCCGCTCGGTTGCGCCCATCCCACCACTCTG |
| 119-L | TTTGTCCGCTCGGTTGCTCAGAGTGGTGTCCTCGG |
| 119-R | TTTGTCCGCTCGGTTGCCGCCCATCCAACCCGAGG |
| link1 | Bio-TGTCCGCTCGGTTGCGGCCGCAAGCTTCCCGGGATCCGCG |
| link2 | CGCGGATCCCGGGAAGCTTGCGGCCGCAACCGAGCGGACA |
| 5S1 | TAYGGGAGAGCGGACGG |
| 5S2 | GTAGTTGGGTCGGTGAC |
| 5SL | NNNACATACGACCATAS |
| 5SR | CAACAGCCGGTATTCGC |

All primers are listed 5' to 3'. The redundant symbols are: Y = C and T at that position; S = C and G at that position; and N = all four nucleotides at that position. Bio- indicates a biotin group covalently attached to the oligonucleotide.

Table 2.3 Thermocycling conditions used for PCR amplification

| Primers involved | PCR thermocycling programme used |
|---|---|
| nts1 - nts2; nts7 - nts8 | 95°C for 2 min; 25 cycles of 95°C for 30 sec, 57°C for 30 sec and 72°C for 1 min; and 72°C for 5 min |
| nts3 - nts4, and reactions involving the M13-forward and reverse primers | 95°C for 2 min; 30 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min; and 72°C for 5 min |
| nts1 - nts8 | 95°C for 2 min; 26 cycles of 95°C for 30 sec, 58°C for 30 sec and 72°C for 2 min; and 72°C for 5 min |
| 5S1 - 5S2 | 95°C for 2 min; 28 cycles of 95°C for 30 sec, 47°C for 30 sec and 72°C for 1 min; and 72°C for 5 min |
| 5SL - 5SR | 95°C for 2 min; 27 cycles of 95°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec; and 72°C for 5 min |

### 2.17.3 STABBED BAND PCR REAMPLIFICATION

Where amplification of only a single product was required in PCR reactions which produced more than one product, stabbed band PCR reamplification was performed. Here the original PCR reaction was fractionated on an agarose mini gel, stained in ethidium bromide and viewed under UV light (Section 2.8.4). The band to be amplified was then stabbed using a 20 - 200 µl pipette tip, and this was suspended in 200 µl of MilliQ $H_2O$. 2 µl of this was used as a template in a subsequent PCR amplification using the same primers (Section 2.17.2). This was repeated as often as necessary to obtain a single PCR product.

### 2.17.4 PURIFICATION OF PCR PRODUCTS

PCR products that required purification (removal of primers, dNTPs, salts and enzymes) for further applications were treated using the PCR Clean Up kit (Boehringer Mannheim) and resuspended in an appropriate volume of MilliQ $H_2O$.

## 2.18 MULTI-VARIANT REPEAT PCR ANALYSIS

A PCR-based technique called multi-variant repeat PCR (MVR-PCR; Jeffreys *et al.*, 1991) was used to map the order of the two variants in the 111/119 bp sub-repeat array (Section 3.1.6) in the Lp1 IGS insert in pPN50 and in Lp1 genomic DNA (Linares *et al.*, 1994). Refer to Section 3.1.8 for details on this technique. To map the 111/119 bp sub-repeats, 18-mer oligonucleotides specific to each of the sub-repeat types were designed in both directions that covered the variable region of the sub-repeats at the 3' end of the primer. In addition an arbitrary 17-mer tail was constructed at the 5' end of each primer to produce the final 35-mer repeat specific primers, and a primer consisting of this arbitrary sequence alone was also constructed. Therefore five primers were created: 111-L and 111-R specific to the 111 bp sub-repeat variant for the left and right hand ends of the sub-repeat array, 119-L and 119-R specific to the 119 bp sub-repeat variant for the left and right hand ends of the sub-repeat array, and the arbitrary TAG primer. These are shown in Table 2.2. The anchor primers, located to the left and right of the 111/119 bp sub-repeat array, are previously designed primers: anchor-L is nts7; and anchor-R is nts4 for genomic DNA and is the M13-forward primer for pPN50 DNA (refer to Table 2.2).

All MVR-PCR reactions consisted of one initial cycle which was carried out in a final volume of 50 µl containing 1 x Expand™ High Fidelity DNA polymerase buffer (Boehringer Mannheim), 5% (v/v) DMSO, 50 µM of each dNTP, 10 nM of the repeat specific primer; 300 nM of the appropriate anchor primer, 3 U of Expand High Fidelity DNA polymerase (Boehringer Mannheim) and 10 ng of Lp1A0 genomic DNA or 0.2 ng of pPN50 DNA. The temperature regime was 3 min at 92°C, 30 sec at 60°C, and 5 min at 70°C. Following this 300 nM of the TAG primer and a further 1 U of Expand High Fidelity DNA polymerase were added to the

reaction, and the reactions were put back into the thermocycler with the following temperature regime: 2 min at 92°C; 25 cycles of 30 sec at 92°C, 30 sec at 60°C, and 3 min at 70°C; and 5 min at 70°C. All reactions were carried out in a Corbett Research thermocycler. Reactions were then separated on 1.5% (w/v) agarose (Section 2.8.1) Horizon mini gels (Section 2.8.2).

## 2.19 CONTOUR-CLAMPED HOMOGENEOUS ELECTRIC FIELD (CHEF) GEL ELECTROPHORESIS

Pulsed field gel electrophoresis was performed using the CHEF system for the separation of very large DNA fragments and intact chromosomes obtained from fungal protoplasts.

### 2.19.1 PREPARATION OF FUNGAL PROTOPLASTS

Protoplasts from various Lp1 cultures, Lp5 and E8 were prepared using a modification of the method described by Murray *et al.* (1992). Fungal cultures were grown in PD broth as described in Section 2.1.2. The mycelia from each culture were pelleted by centrifugation at 27,000 g for 15 min at 4°C, and resuspended in an equal volume of sterilised MilliQ $H_2O$ and pelleted three times. 20 ml of OM buffer (Section 2.2.4) containing 20 mg/ml of Novozyme (Novo Industri A/S) or Glucanex (Chemcolour Industry NZ), depending on availability, were filter-sterilised by passage through a 45 µm Supor® Acrodisk® 32 (Gelman Sciences) filter, and the mycelium was resuspended in this. The mixture was incubated at 28°C at 200 rpm for 3 - 5 h until protoplasting was complete (judged by the fungal mass reducing to a milky solution). The protoplasts were separated from the undigested hyphae by filtering the mixture through mira-cloth (Calbiochem) and the filtrate was overlaid with 2 ml of ST buffer (Section 2.2.4). The protoplasts were banded at the interface by centrifugation at 3,000 *g* for 5 min at 4°C. After removal from the interface, the protoplasts were washed three times in 5 ml of STC buffer (Section 2.2.4) and pelleted each time by centrifugation at 7,700 *g* for 5 min at 4°C. The protoplasts were then resuspended in GMB buffer (Section 2.2.4) to give a final concentration of 1 x $10^9$ protoplasts/ml, as determined by haemocytometer counting under a microscope.

### 2.19.2 PREPARATION OF DNA IN AGAROSE-EMBEDDED PLUGS

The protoplasts prepared from the previous section were mixed with an equal volume of LMP in GMB (Section 2.2.4), immediately pipetted into CHEF plug molds (Bio-Rad) and allowed to set for 20 min at 4°C. The plugs were then incubated in 10 ml of SE buffer (Section 2.2.4) at 55°C for 20 h, followed by incubation at 50°C for 24 h in 10 ml of 10 x ET buffer (Section 2.2.4) supplemented with 20 mg of proteinase K (Boehringer Mannheim) and 0.1 g of sodium lauroylsarcosine (Sigma). Finally the plugs were washed four times for at least 3 h each with 10 ml of 1 x ET buffer (Section 2.2.4) and stored at 4°C.

### 2.19.3 RESTRICTION ENDONUCLEASE DIGESTION OF AGAROSE-EMBEDDED PLUGS

Digestion of DNA in agarose-embedded plugs by restriction endonucleases for separation by CHEF gel electrophoresis was performed as follows. Two plugs prepared as outlined in the previous section were washed twice in 10 ml of TE (10/0.1) buffer (Section 2.2.4) for 2 h each at 37°C with gentle agitation, followed by another two washes with the same conditions for 1 h each. The plugs were transferred into a 250 µl pre-reaction mixture containing the commercial restriction endonuclease buffer recommended by the manufacturer and acetylated bovine serum albumin (Promega) to a final concentration of 100 µg/ml, and were incubated at 37°C for 30 min. The plugs were then transferred to an identical, fresh reaction mixture, 500 U of the restriction endonuclease were added and this was incubated for 16 h at 37°C. The digested plugs were stored at 4°C.

### 2.19.4 CHEF GEL ELECTROPHORESIS CONDITIONS

All CHEF gel electrophoresis was performed on a BioRad CHEF-DRII™ apparatus in 0.5 x TBE buffer (Section 2.2.4) at 12°C. CHEF gels 140 x 125 mm in size were made using chromosomal grade agarose (Bio-Rad) at the concentration specified in each figure, in 100 ml of 0.5 x TBE buffer. Plugs were initially checked for concentration and quality on a 1.0% (w/v) CHEF gel using the following conditions: a ramped switch time of 100 - 1,000 sec for 24 h at 100 volts. Two sets of conditions for full separations were used. For separation of chromosome-sized bands (undigested plugs and plugs digested with the intron-encoded endonuclease, I-*Ppo*I [Promega]) the following conditions were used: a switch time of 120 sec for 15 h, followed by a switch time of 450 sec for 13 h, then a switch time of 1,600 sec for 35 h and a switch time of 2,000 sec for 35 h, all at 60 volts. For separation of rDNA clusters (plugs digested with *Bam*HI or *Hin*dIII) the following conditions were used: a switch time ramped from 120 sec to 150 sec over 24 h at 150 volts. Detection of DNA was performed as described in Section 2.8.4.

## 2.20 PCR-BASED CLONING OF RDNA JUNCTION FRAGMENTS

### 2.20.1 PREPARATION OF GENOMIC (OR VECTOR) DNA

6 µg of Lp5 or E8 genomic DNA (or pUC118 DNA) was digested with *Hin*dIII or *Bam*HI (Section 2.5). The high molecular weight (greater than 20 kb) genomic DNA (but not the plasmid DNA) was gel extracted from a Seaplaque agarose gel (Section 2.9) and resuspended in 20 µl of MilliQ $H_2O$. The plasmid DNA was phenol/chloroform extracted (Section 2.6) ethanol precipitated and resuspended in 20 µl of MilliQ $H_2O$.

### 2.20.2 PREPARATION OF LINKER DNA

5 µg of each complementary linker oligonucleotide (Table 2.2) were annealed in 1 x annealing solution (final concentration; Section 2.2.4) by heating to 65°C for 10 mins and allowing this to cool at room temperature. This mixture was digested to completion with either *Hin*dIII or *Bam*HI (Section 2.5). The digested linker was phosphatase treated by addition of 10 µl of 10 x dephosphorylation buffer and 5 µl of shrimp alkaline phosphatase (Boehringer Mannheim) in a total volume of 100 µl, and this was incubated for 90 min at 37°C before inactivation of the enzyme by heating at 65°C for 15 min. The linker was phenol/chloroform extracted (Section 2.6), ethanol precipitated (Section 2.7) and resuspended in 20 µl of MilliQ $H_2O$.

### 2.20.3 PREPARATION OF DYNABEADS®

Magnetic bead technology was used in the enrichment of junction fragments (Section 2.20.4). The streptavidin-coated magnetic Dynabeads® (Dynal) were prepared as follows: 100 µl of the Dynabeads were dispensed into an Eppendorf tube, placed on a MagneSphere® Technology separation stand (Promega) and the supernatant removed with an autopipetter. The beads were washed three times in 100 µl of PBS/BSA (Section 2.2.4), with the wash solution removed each time on the magnetic bead separator with an autopipetter. They were then washed once with 100 µl of 1 x DBW buffer (Section 2.2.4), and were finally resuspended in 200 µl of 2 x DBW buffer.

### 2.20.4 ENRICHMENT OF JUNCTION FRAGMENTS

Ligation was performed on the DNA samples from Sections 2.20.1 and 2.20.2 by mixing 10 µl of the gel extracted genomic DNA (or 5 µl of plasmid DNA) with 2 µl of the linker DNA in the presence of 200 U of T4 DNA ligase and a final concentration of 1 x T4 DNA ligase buffer (both New England Biolabs) in a total volume of 20 µl. This gives a large molar excess of linker DNA over genomic DNA. After overnight incubation at 4°C, this was phenol/chloroform extracted (Section 2.6), ethanol precipitated (Section 2.7) and the genomic DNA samples were digested with the intron-encoded endonuclease, I-*Ppo*I (Promega) according to manufacturer's instructions. The solution was made up to 200 µl with MilliQ $H_2O$ and was mixed with 200 µl of the prepared Dynabeads (Section 2.20.3). This mixture was gently agitated for at least 30 min at room temperature to allow the biotin-labelled linker DNA to bind to the streptavidin-coated Dynabeads. The beads were then washed three times in 500 µl of 1 x DBW buffer (Section 2.2.4) and the wash solution removed each time with a magnetic bead separator and autopipetter. Finally the beads were resuspended in 100 µl of MilliQ $H_2O$.

### 2.20.5 PCR AMPLIFICATION OF JUNCTION FRAGMENTS

For plasmid DNA, PCR amplification with the M13-forward and M13-reverse primers and the TAG primer (Table 2.2) was performed as set out in Section 2.17. For genomic DNA the first round of PCR amplification was performed using primers nts1, nts2, nts3 or nts4 with the TAG primer (Table 2.2) in 50 µl reaction volumes. The reaction conditions were as set out in Section 2.17.2, except 2 U of Expand™ Long Template DNA polymerase (Boehringer Mannheim) and 1.3 mM of each dNTP was used, and the template DNA was 8 µl of the Dynabead suspension (Section 2.20.4). These reactions were placed in a Corbett Research thermocycler, where the following thermocycling conditions were employed: 92°C for 2 min; 10 cycles of 92°C for 10 sec, 58°C for 30 sec and 68°C for 12 min; 15 cycles of 92°C for 10 sec, 58°C for 30 sec and 68°C for 12 min plus an additional 20 sec/cycle; and 68°C for 7 min. These reactions were then run on an agarose mini gel (Section 2.8.2).

The second round of PCR amplification was performed using the nested primers nts11, nts12, nts13 or nts14 again with the TAG primer (Table 2.2). The reaction and thermocycling conditions were exactly the same as the first round of amplification, except that 5 µl of a 500-fold dilution of PCR reaction from the first round of amplification was used as the template. These reactions were then run on an agarose mini gel.

Any bands amplified from the nested round of PCR amplification were individually amplified using stabbed band PCR reamplification (Section 2.17.3) with the same primers as used in the nested PCR amplification. The PCR was performed in 50 µl reaction volumes with the PCR conditions as set out in Section 2.17.2, except that Expand™ High Fidelity DNA polymerase (Boehringer Mannheim) was always used. These reactions were placed in a Corbett Research thermocycler, where the following thermocycling conditions were employed: 92°C for 2 min; 30 cycles of 92°C for 10 sec, 60°C for 30 sec and 68°C for 7 min; and 68°C for 10 min. The reactions were then run on an agarose mini gel.

### 2.20.6 CLONING OF POTENTIAL JUNCTION FRAGMENTS

Potential junction fragments that were individually amplified (Section 2.20.5) were cloned into pGEM®-T for further analysis. The PCR products were purified according to Section 2.17.4. These PCR products were A-tailed to ensure efficient ligation into the pGEM-T vector by incubation in a 10 µl solution containing 1 x *Taq* DNA polymerase buffer, 5 U of *Taq* DNA polymerase, 0.2 mM dATP (all Boehringer Mannheim) and 2.5 mM $MgCl_2$ for 30 min at 70°C. These A-tailed PCR products were then ligated into the pGEM-T vector (Section 2.12.2) and transformants containing the correct insert were selected (Section 2.13).

# CHAPTER THREE

# RESULTS

# 3.1 CHARACTERISATION OF EXTRAORDINARY RIBOSOMAL IGS LENGTH HETEROGENEITY IN THE HYBRID, LP1

Previous work had shown there was an extraordinary amount of length variation present in the ribosomal DNA in the hybrid endophyte, Lp1 (Ganley, 1993). I was interested in characterising this IGS length heterogeneity and identifying its cause, as I believed this related to the concerted evolution of the rDNA.

## 3.1.1 GENERATION OF SINGLE-SPORE ISOLATES

To facilitate characterisation of the ribosomal length heterogeneity I generated a series of single-spore isolates from Lp1. Asexual isolates, including the interspecific hybrids, retain the ability to produce conidiospores. These spores are uninucleate (Schardl *et al.*, 1994), allowing pure cultures of a homogenous nuclear composition to be generated from a mycelium by the germination of a Single-spore. The original Lp1 isolate (called Lp1O in this study) had previously been split into four laboratory cultures that were maintained separately for one to four years before the commencement of this study. These cultures are indicated by the first letter following the Lp1 designation (e.g. Lp1A; see Table 2.1 for the origins of the different laboratory cultures). Two single-spore isolates were obtained from each of these laboratory cultures, and these are designated by a number following the laboratory culture letter designation (e.g. Lp1A3). Two single-spore isolates were then obtained from each of these first-round single-spore isolates (except from Lp1F3 and Lp1F6), to give a series of second-round single-spore isolates. These are designated by a letter following the first-round single-spore number designation (e.g. Lp1A3b). A similar set of single-spore isolates was generated from the independently-isolated LpTG-2 hybrid isolate, Lp2, which was designated Lp2G.

## 3.1.2 LENGTH HETEROGENEITY OCCURS WITHIN THE rDNA CLUSTER AND IS STILL BEING GENERATED

To ascertain whether the extraordinary variation in rDNA length observed by Ganley (1993) is the result of intercellular differences between rDNA clusters or occurs within an rDNA cluster, genomic DNA was extracted from each of the laboratory Lp1 cultures and the Lp2 culture, as well as from all the single-spore purified isolates. The DNA from these laboratory cultures was digested with *Sal*I and separated on agarose gels (Figure 3.1). There are no discernible differences in the ethidium bromide-stained banding profiles for any of these cultures.

The gels from Figure 3.1 were Southern blotted and probed with the insert from pPN50, containing the IGS region on a 4.1 kb *Sal*I fragment (referred to throughout as the 4.1 kb *Sal*I IGS probe). Figures 3.2A and 3.2B show the original laboratory cultures with their two first-round single-spore isolates alongside. Figures 3.2C, 3.2D, and 3.2E show the first-round

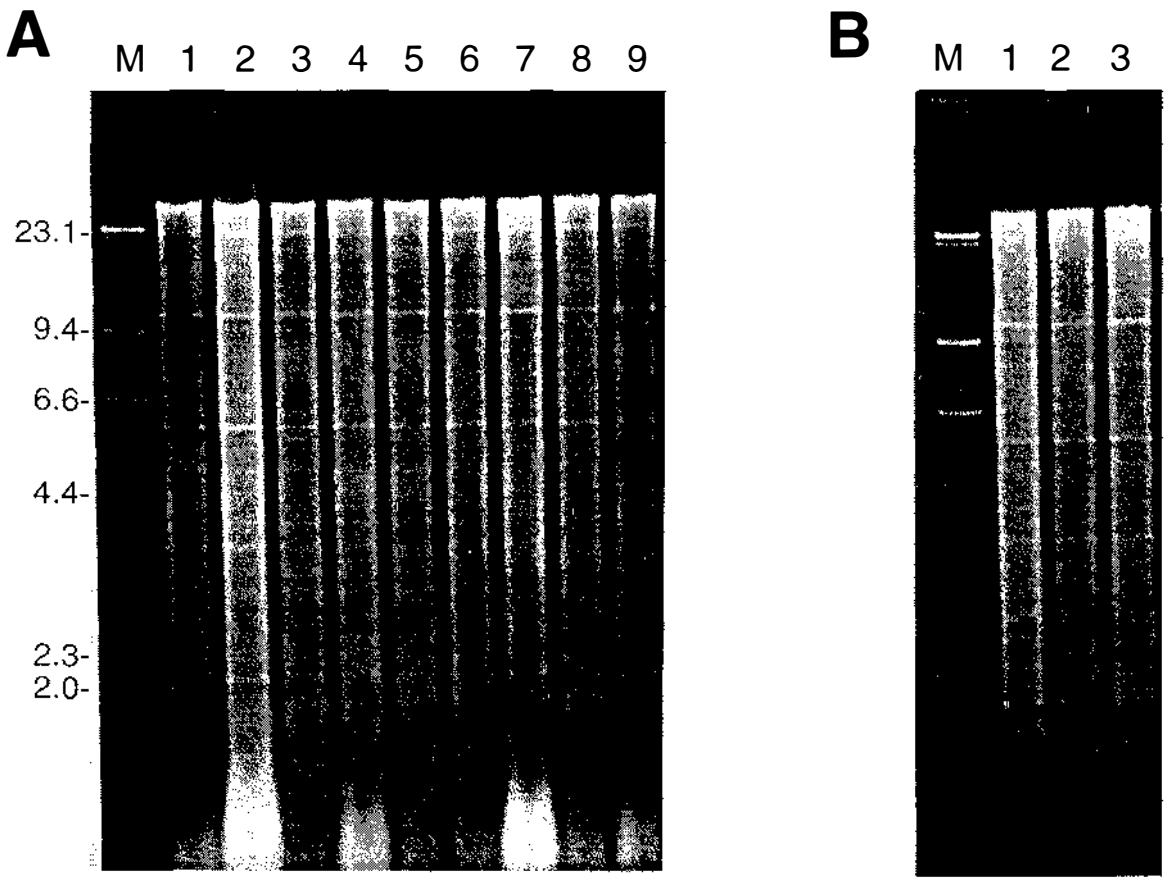



Figure 3.1 *Sal*I genomic digests of Lp1 laboratory cultures and single-spore purified isolates

*Sal*I digests were performed on 1.2 μg of genomic DNA. These were fractionated on 0.7% agarose gels and the gels stained with ethidium bromide. In panels A, B and C lane M corresponds to the λ *Hin*dIII size marker, and the sizes are shown in kb to the left of panel A.

A Lp1A0 (lane 1), Lp1A3 (lane 2), Lp1A6 (lane 3), Lp1C0 (lane 4), Lp1C3 (lane 5), Lp1 C6 (lane 6), Lp1D0 (lane 7), Lp1D3 (lane 8) and Lp1D6 (lane 9).

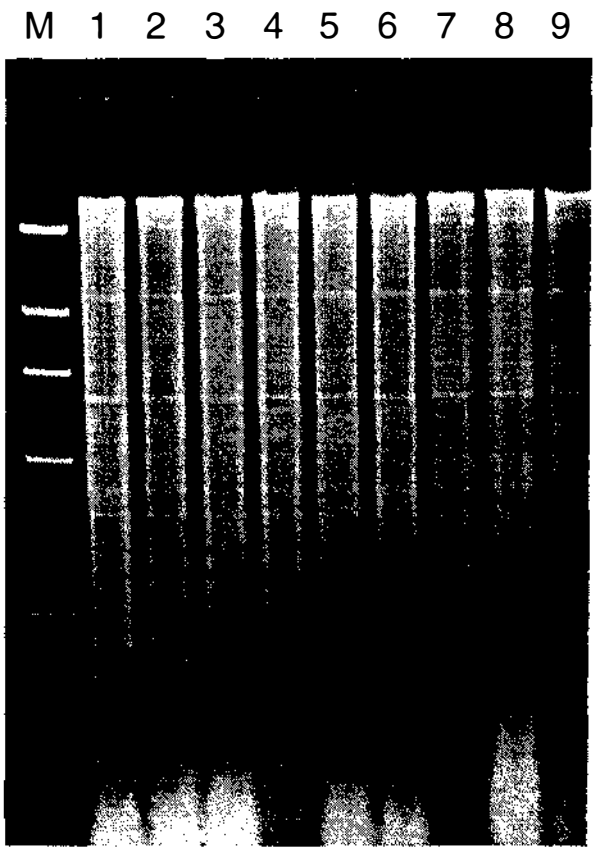
B Lp2G0 (lane 1), Lp2G3 (lane 2) and Lp2G6 (lane 3).

C Lp1A3 (lane 1), Lp1A3b (lane 2), Lp1A3d (lane 3), Lp1A6 (lane 4), Lp1A6b (lane 5), Lp1A6d (lane 6), Lp1C3 (lane 7), Lp1C3b (lane 8) and Lp1C3d (lane 9).

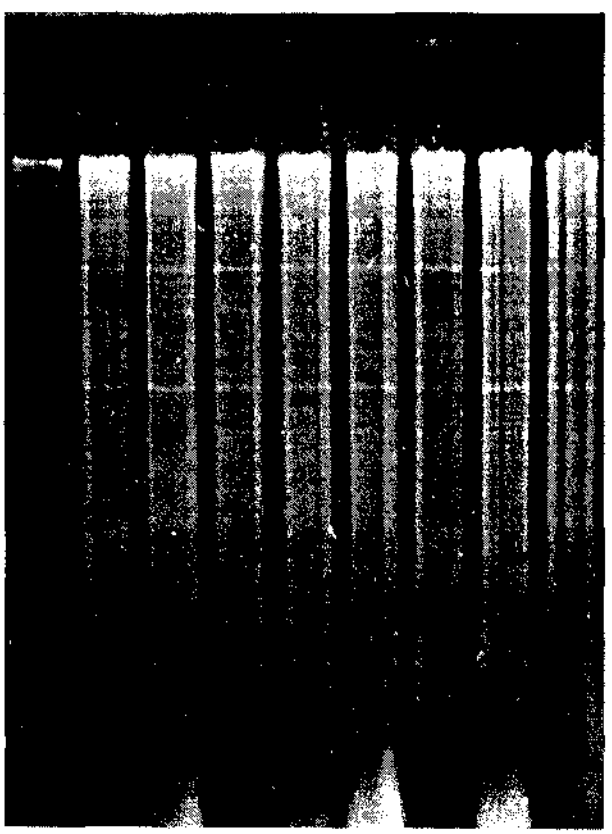
D Lp1C6 (lane 1), Lp1C6b (lane 2), Lp1C6d (lane 3), Lp1D3 (lane 4), Lp1D3b (lane 5), Lp1D3d (lane 6), Lp1D6 (lane 7), Lp1D6b (lane 8) and Lp1D6d (lane 9).

E Lp2G3 (lane 1), Lp2G3b (lane 2), Lp2G3d (lane 3), Lp2G6 (lane 4), Lp2G6b (lane 5) and Lp2G6d (lane 6).

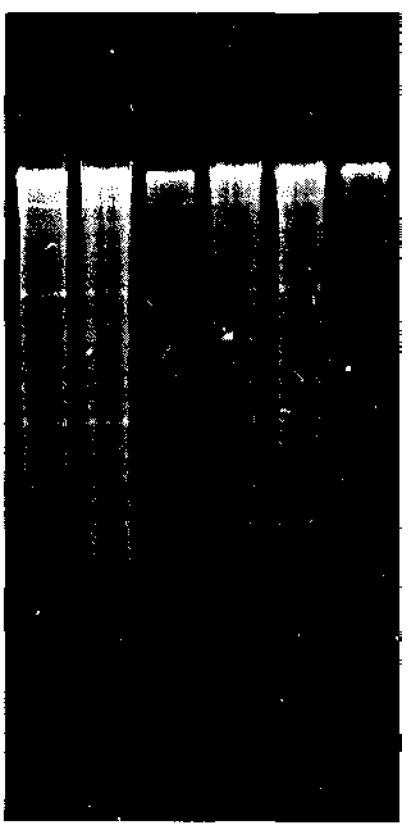
C




D




E



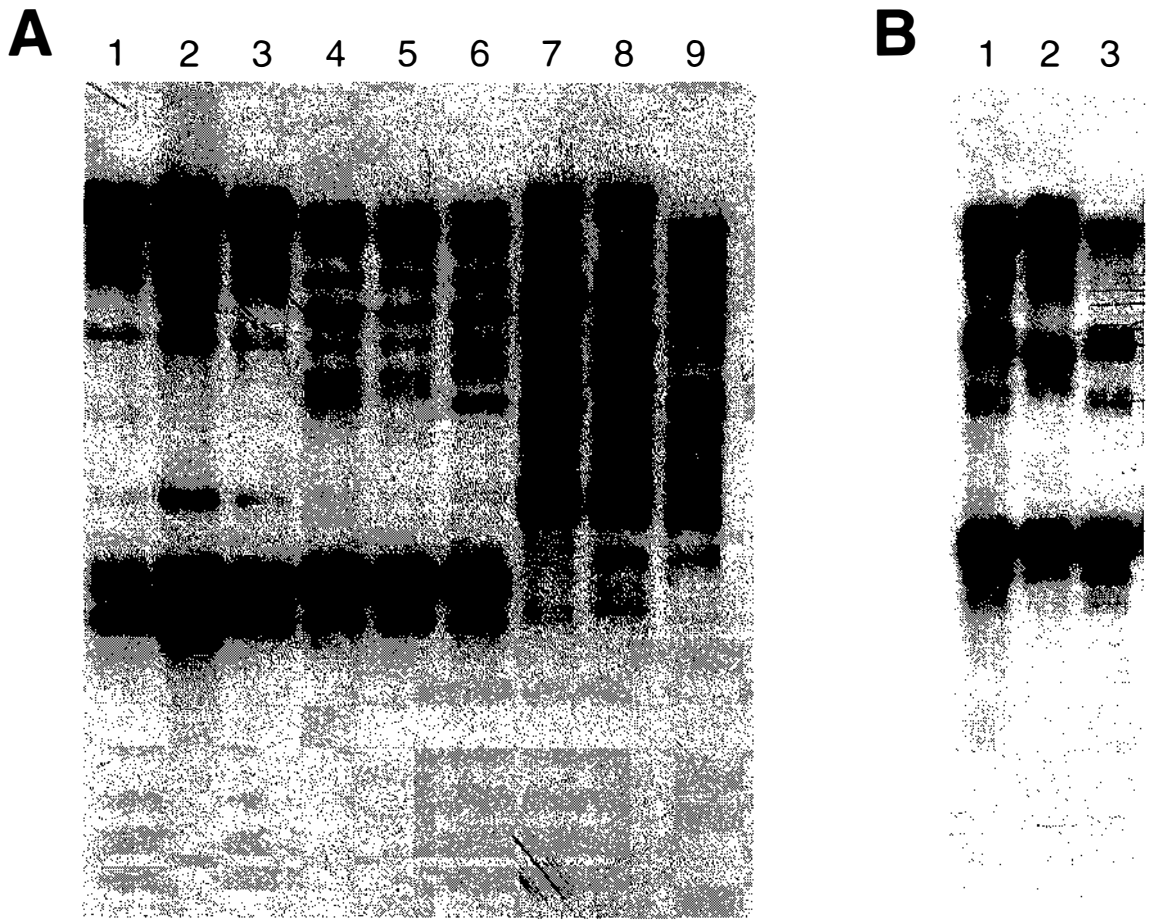



Figure 3.2 Southern analysis of the *Sal*I Lp1 genomic digests from Figure 3.1 with the 4.1 kb *Sal*I IGS probe

The respective gels shown in Figure 3.1 were Southern blotted and probed with the 4.1 kb *Sal*I IGS probe. Examples of hybridising bands that represent several rDNA units that are either lost or gained through the single-sporing are indicated by arrows. In panels C and D lane M corresponds to the λ *Hin*dIII size marker and the sizes are shown in kb to the left of panel C.

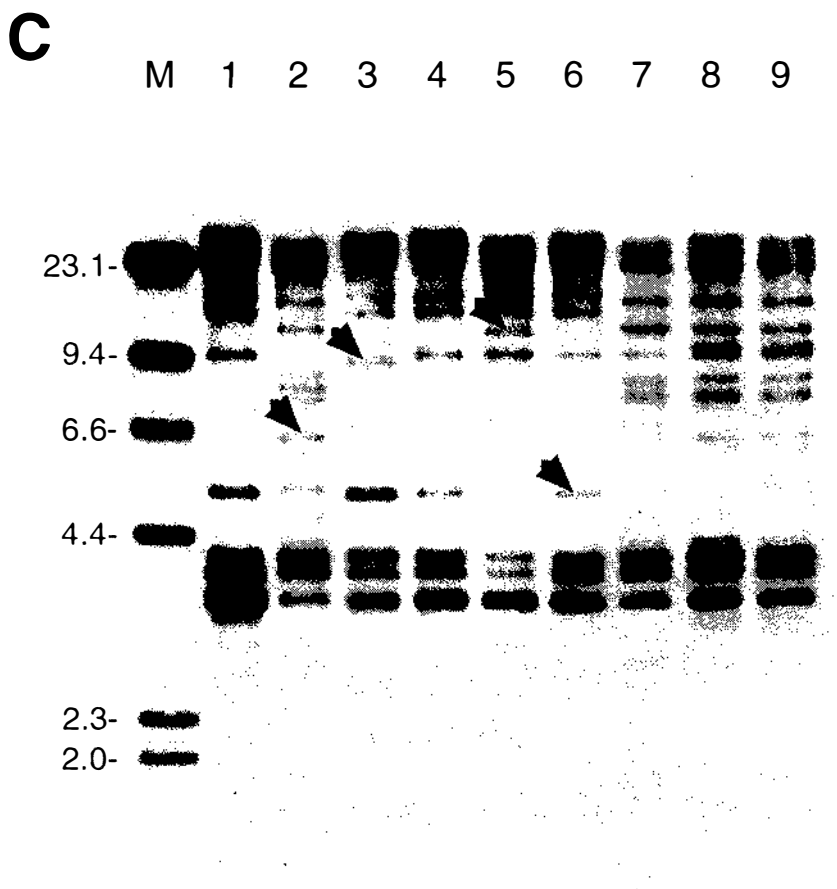
A Lp1A0 (lane 1), Lp1A3 (lane 2), Lp1A6 (lane 3), Lp1C0 (lane 4), Lp1C3 (lane 5), Lp1 C6 (lane 6), Lp1D0 (lane 7), Lp1D3 (lane 8) and Lp1D6 (lane 9).

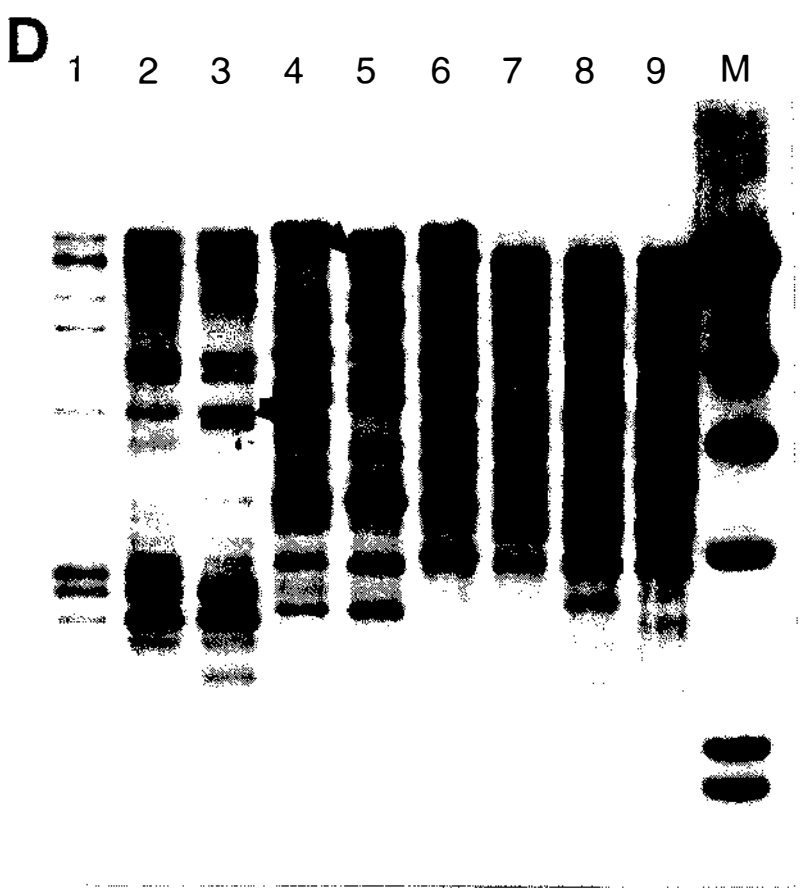
B Lp2G0 (lane 1), Lp2G3 (lane 2) and Lp2G6 (lane 3).

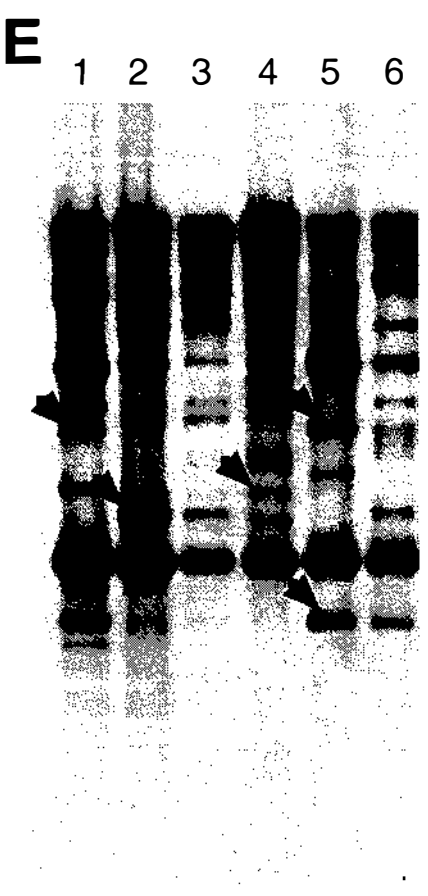
C Lp1A3 (lane 1), Lp1A3b (lane 2), Lp1A3d (lane 3), Lp1A6 (lane 4), Lp1A6b (lane 5), Lp1A6d (lane 6), Lp1C3 (lane 7), Lp1C3b (lane 8) and Lp1C3d (lane 9).

D Lp1C6 (lane 1), Lp1C6b (lane 2), Lp1C6d (lane 3), Lp1D3 (lane 4), Lp1D3b (lane 5), Lp1D3d (lane 6), Lp1D6 (lane 7), Lp1D6b (lane 8) and Lp1D6d (lane 9).

E Lp2G3 (lane 1), Lp2G3b (lane 2), Lp2G3d (lane 3), Lp2G6 (lane 4), Lp2G6b (lane 5) and Lp2G6d (lane 6).


C
M 1 2 3 4 5 6 7 8 9
23.1-
9.4-
6.6-
4.4-
2.3-
2.0-


D
1 2 3 4 5 6 7 8 9 M


E
1 2 3 4 5 6

single-spore isolates with their two second-round single-spore isolates alongside. A multitude of bands ranging in size from 3.5 kb to greater than 20 kb hybridise for these cultures, including a band the size of the insert in pPN50 (4.1 kb). Each laboratory culture retained its distinctive banding pattern through the single-sporing rounds with the 4.1 kb *Sal*I IGS probe. This demonstrates that the rDNA length variants are present within the rDNA cluster - if they resulted from intercellular rDNA differences only a subset of the bands seen in the original laboratory cultures would be expected in the single-spore isolates.

While the banding patterns with the 4.1 kb IGS probe remained broadly similar through the single-sporing, some changes were observed. This indicates that the mechanism (s) causing the rDNA length heterogeneity is still active. Interestingly, in many cases these changes involved strongly-hybridising bands either appearing or disappearing in one round of single-spore isolation (indicated in Figure 3.2). This suggests these bands consist of a number of length variants that are clustered together in the array.

The blots were stripped and reprobed with the insert from pPN49, containing the three ribosomal RNA (*rrn*) genes on a 5.6 kb *Sal*I fragment (referred to throughout as the 5.6 kb *Sal*I coding region probe; Figure 3.3). No evidence of the extraordinary level of length heterogeneity seen in Figure 3.2 is seen with this probe. Instead there is a single strongly-hybridising band, 5.6 kb in size, corresponding exactly to that expected from the initial mapping (Ganley 1993; Figure 1.1). This demonstrates that the pattern of bands seen with the 4.1 kb *Sal*I IGS probe cannot be the result of partial digestion by *Sal*I. The hybridisation signal in Figure 3.2 is proportional to the DNA loading seen in Figure 3.1.

### 3.1.3 LENGTH VARIATION OF THE IGS IS NOT OBSERVED IN EITHER PROGENITOR OF Lp1

Genomic DNA from the two putative progenitors of Lp1, *Neotyphodium lolii* LpTG-1 isolate Lp5 and *Epichloë typhina* isolate E8, was digested with *Sal*I and a Southern blot probed with the 4.1 kb *Sal*I IGS and 5.6 kb *Sal*I coding region probes (Figure 3.4). Lp1 (single-spore isolate Lp1D3b) is shown alongside for comparison, as is the λPN1 from which pPN49 and pPN50 were derived. The extraordinary level of length heterogeneity seen in Lp1 with the 4.1 kb *Sal*I IGS probe is not present in either progenitor, although more than one band is observed in these two isolates. In Lp5 three bands, ranging from 3.5 kb to 3.7 kb, hybridise to the 4.1 kb *Sal*I IGS probe, as well as weak bands at 4.1 and 5.2 kb. This limited length heterogeneity is qualitatively different to that seen with Lp1. In the case of E8 the only hybridising bands are a doublet around 3.8 kb, and some very weak high molecular weight bands around 9.4 kb and > 20 kb. Aside from some faint hybridisation around 3.2 kb in Lp1, no hybridising bands smaller than those in the two progenitors are seen. Thus the bands seen in Lp1 are at least as great in size as the bands present in the two progenitors. In the λ clone a 4.1 kb band hybridises as expected.

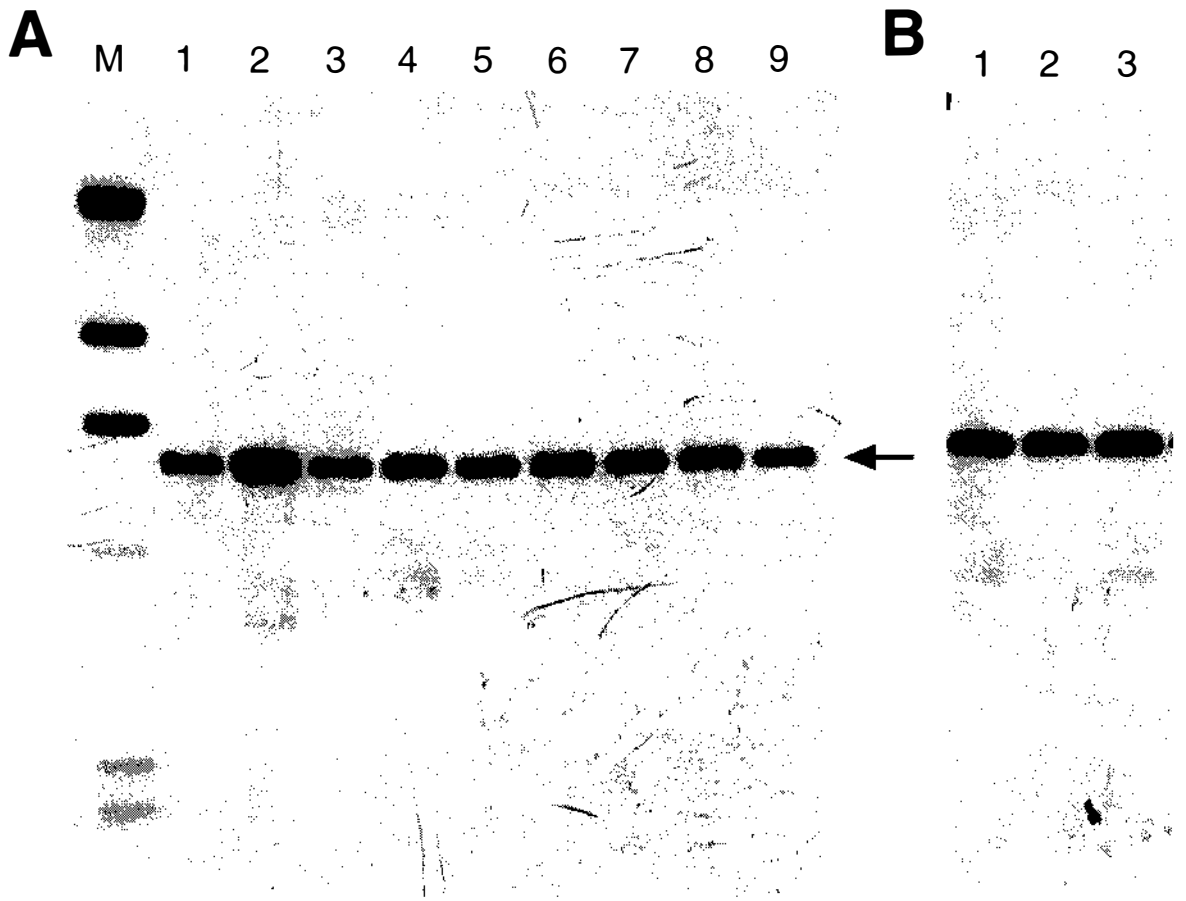



Figure 3.3 Southern analysis of the *Sal*I Lp1 genomic digests from Figure 3.1 with the 5.6 kb *Sal*I coding region probe

The respective Southern blots from Figure 3.2 were stripped and reprobed with the 5.6 kb *Sal*I coding region probe. Lane M corresponds to the λ *Hin*dIII size marker in panel A, and the 5.6 kb hybridising band present in all of these isolates is indicated by an arrow in panel A.

A Lp1A0 (lane 1), Lp1A3 (lane 2), Lp1A6 (lane 3), Lp1C0 (lane 4), Lp1C3 (lane 5), Lp1 C6 (lane 6), Lp1D0 (lane 7), Lp1D3 (lane 8) and Lp1D6 (lane 9).

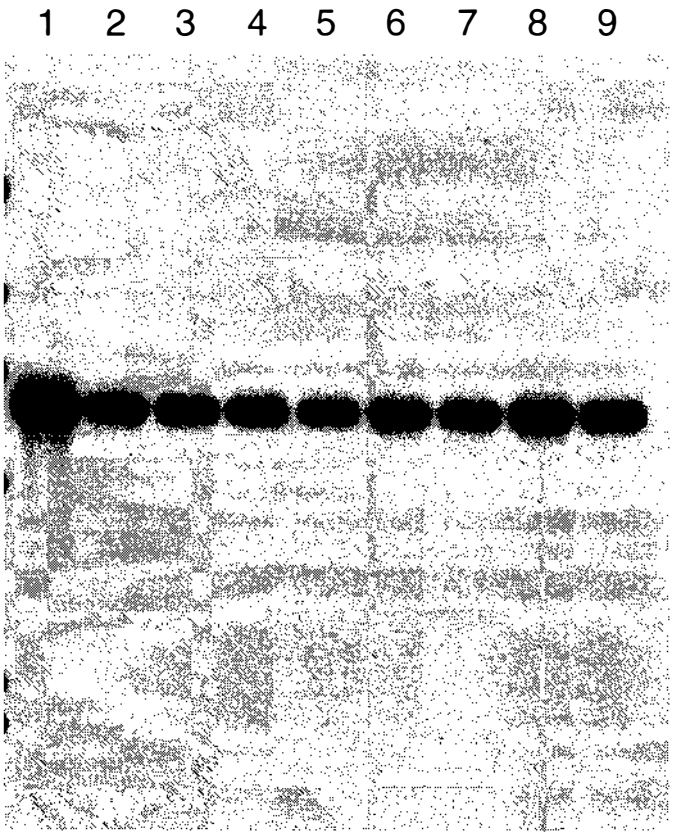
B Lp2G0 (lane 1), Lp2G3 (lane 2) and Lp2G6 (lane 3).

C Lp1A3 (lane 1), Lp1A3b (lane 2), Lp1A3d (lane 3), Lp1A6 (lane 4), Lp1A6b (lane 5), Lp1A6d (lane 6), Lp1C3 (lane 7), Lp1C3b (lane 8) and Lp1C3d (lane 9).

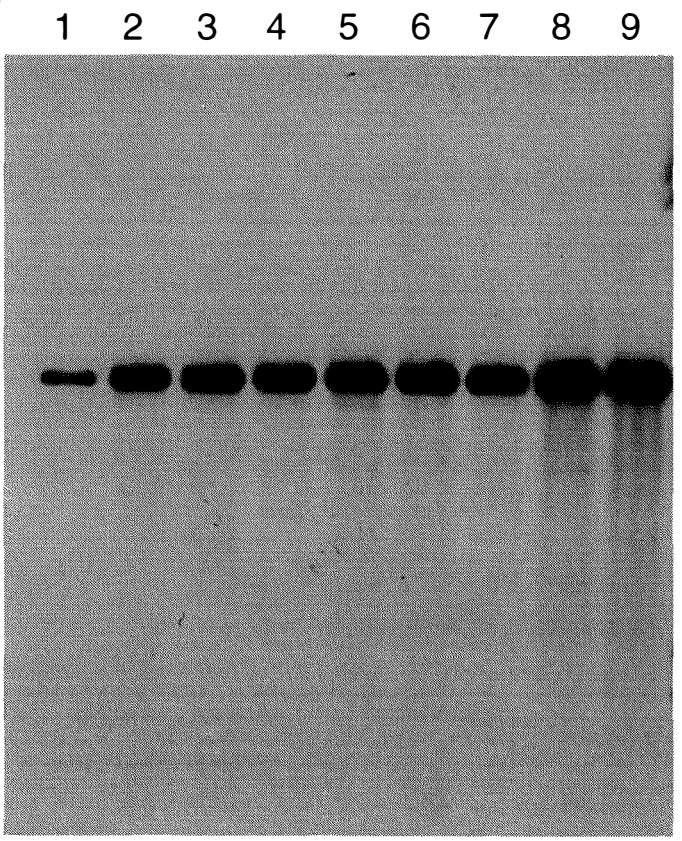
D Lp1C6 (lane 1), Lp1C6b (lane 2), Lp1C6d (lane 3), Lp1D3 (lane 4), Lp1D3b (lane 5), Lp1D3d (lane 6), Lp1D6 (lane 7), Lp1D6b (lane 8) and Lp1D6d (lane 9).

E Lp2G3 (lane 1), Lp2G3b (lane 2), Lp2G3d (lane 3), Lp2G6 (lane 4), Lp2G6b (lane 5) and Lp2G6d (lane 6).

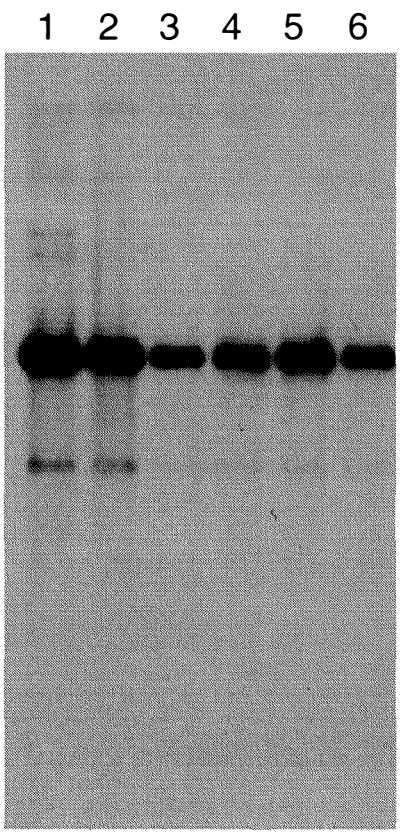
C




D




E



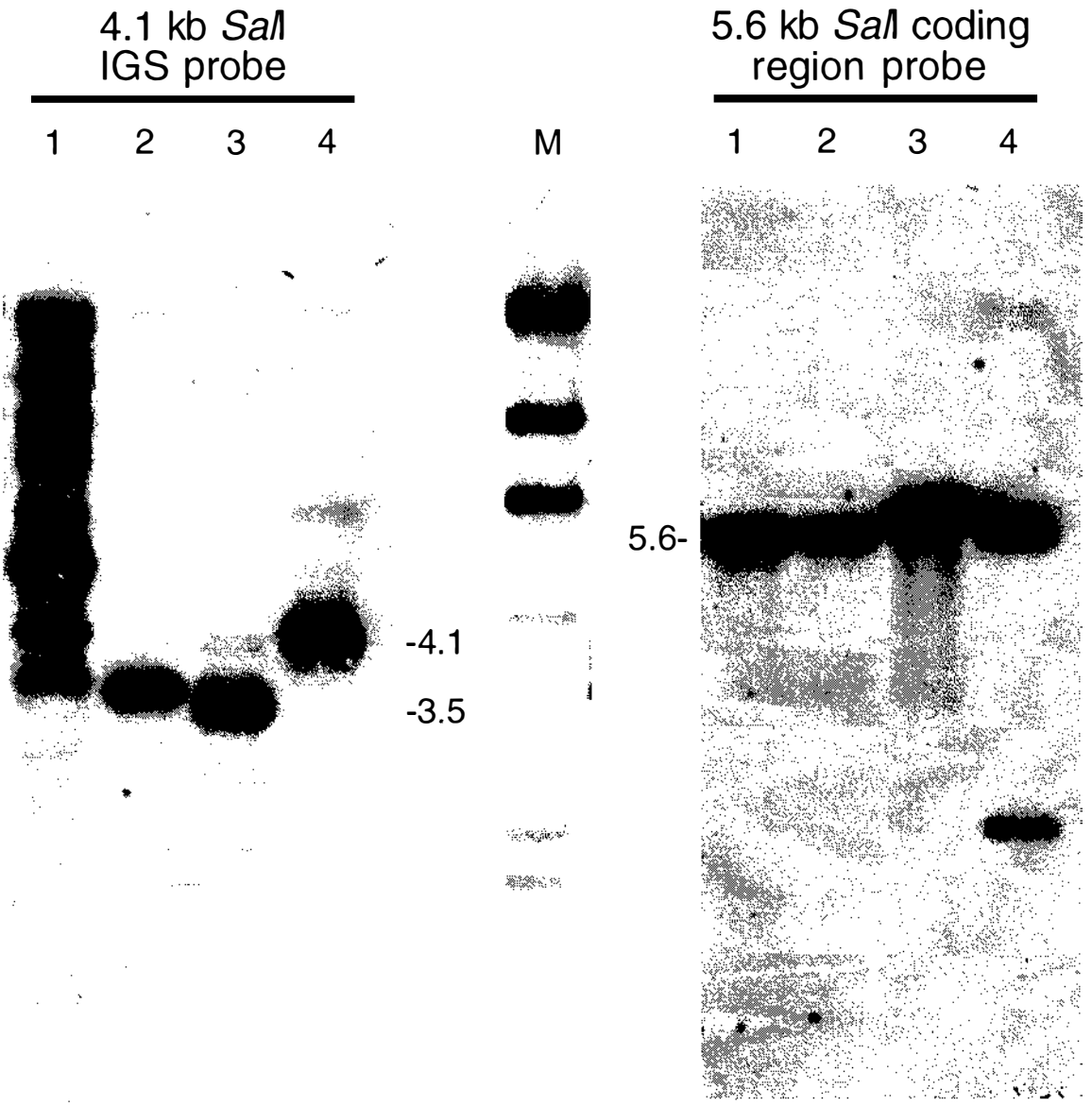



Figure 3.4 The progenitors of Lp1 do not display ribosomal length heterogeneity

1.2 µg of genomic DNA from Lp1D3b (lane 1), E8 (lane 2) and Lp5 (lane 3), and 20 ng of λPN1 DNA (lane 4) was digested with *Sal*I and fractionated on a 0.7% agarose gel. The resulting Southern blot was probed with the 4.1 kb *Sal*I IGS probe, and then stripped and reprobed with the 5.6 kb *Sal*I coding region probe as indicated. The smallest IGS length in Lp1 (3.5 kb), the IGS length equivalent to the clone (4.1 kb) and the 5.6 kb coding region band are all indicated. Lane M is the λ *Hin*dIII size marker.

When these Southern blots are probed with the 5.6 kb *Sal*I coding region probe there is a single hybridising band. In Lp5 this band is approximately 500 bp larger than Lp1 and E8, being around 6.1 kb in size. The weakly-hybridising band in the λ clone (lane 4) is the 2.2 kb *Sal*I truncated coding region fragment (refer to Figure 1.1).

### 3.1.4 LENGTH HETEROGENEITY LOCALISED TO THE INTERGENIC SPACER

To rule out the possibility that the banding pattern seen in Figure 3.2 resulted from heterologous hybridisation of the 4.1 kb IGS probe, genomic DNA from two representative Lp1 single-spore purified cultures (Lp1A3 and Lp1C3), and from Lp5 and E8 was cleaved with *Eco*RI, Southern blotted and probed with the 4.1 kb *Sal*I IGS and 5.6 kb *Sal*I coding region probes (Figure 3.5). The physical map of λPN1 (Figure 1.1) shows that an *Eco*RI genomic digest should produce two rDNA fragments, one corresponding approximately to the 5.8S and 28S *rrn* genes (3.2 kb in size), and the other corresponding to the 18S *rrn* gene and the IGS (6.5 kb in size with the 4.1 kb IGS).

Probing with the 4.1 kb *Sal*I IGS probe produced the same general heterogeneous banding pattern seen in the *Sal*I digestion, with the bands all greater in size by 2.5 kb (the size of the 18S *rrn* gene) than the bands seen in Figure 3.2. The 5.6 kb *Sal*I coding region probe hybridised to a 3.2 kb band as expected (except for Lp5, where this band was ~ 500 bp larger, consistent with Figure 3.4), and also hybridised to the same multitude of bands that the 4.1 kb *Sal*I IGS probe hybridised to. The heterogeneous banding pattern seen with this probe is the result of linkage of the 18S *rrn* gene and the IGS in the *Eco*RI digest, with the probe hybridising to the 18S moiety.

These results demonstrate that the heterogeneous banding pattern is not the result of heterologous hybridisation of the 4.1 kb *Sal*I IGS probe, but is the result of heterogeneity in the IGS region of the rDNA. They also confirm that the *Sal*I bands seen in Figures 3.2 and 3.3 are linked *in vivo*, and that the multitude of bands are ribosomal in origin.

The same pattern of hybridisation is seen with the two progenitors, E8 and Lp5, but without the multitude of hybridising bands as expected. The same weakly-hybridising bands seen with the 4.1 kb *Sal*I IGS probe in Figure 3.4 can be detected with both probes, although there is an additional weak band in E8 around 8.3 kb with the 4.1 kb *Sal*I IGS probe (Figure 3.5, lane 3).

These results also rule out the possibility that the hybridisation pattern is the result of an unlikely experimental artefact, such as 'star' activity of the restriction endonucleases or methylation of the rDNA, as the same pattern of results for the two restriction endonucleases used would not be expected under either scenario.

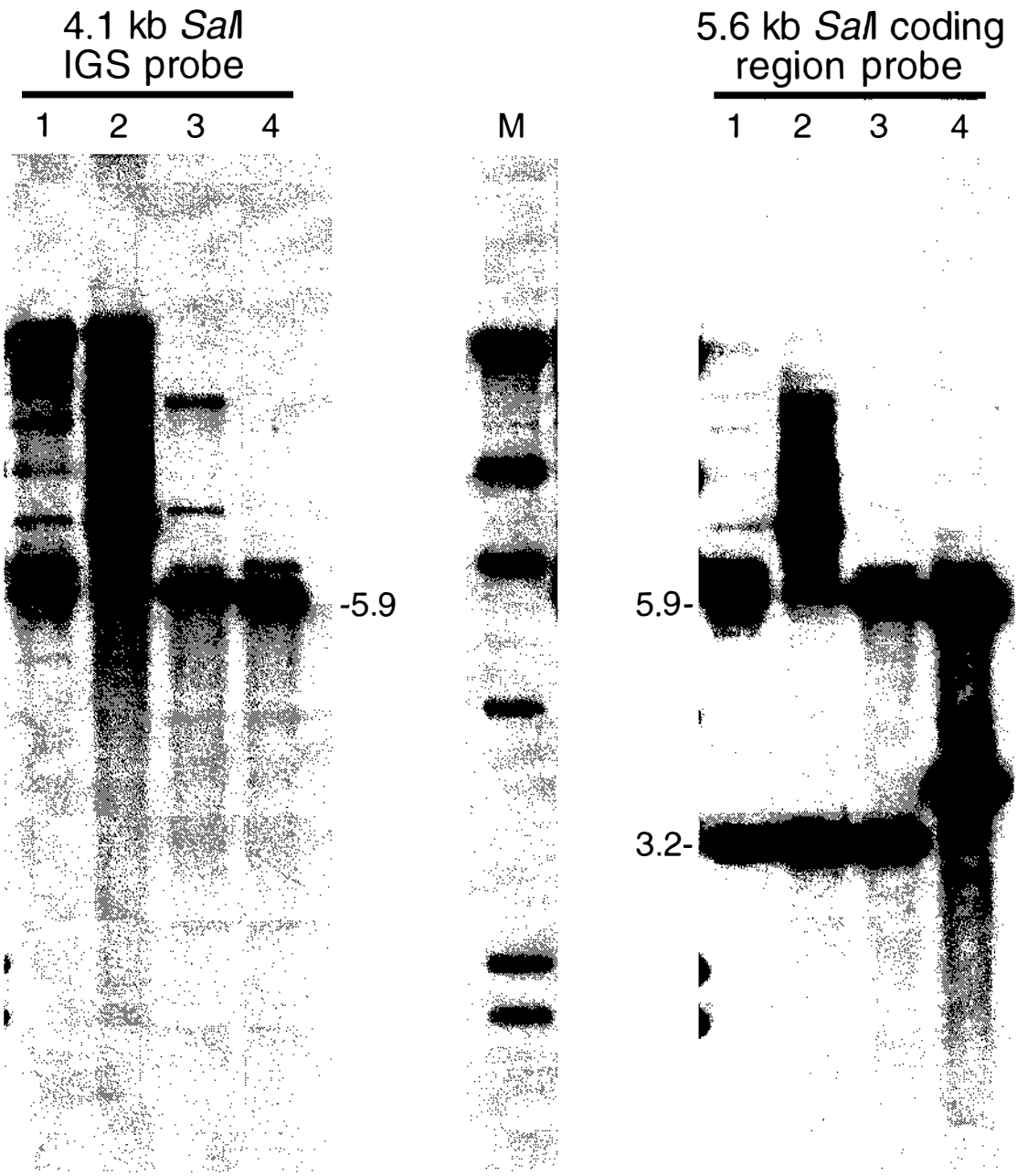



Figure 3.5 Length heterogeneity is specific to the Lp1 ribosomal IGS region

1.2 µg of genomic DNA from Lp1A3 (lane 1), Lp1C3 (lane 2), E8 (lane 3) and Lp5 (lane 4) was digested with *Eco*RI and fractionated on a 0.7% agarose gel. The resulting Southern blot was probed with the 4.1 kb *Sal*I IGS probe, and then stripped and reprobed with the 5.6 kb *Sal*I coding region probe as indicated. The smallest IGS length (5.9 kb), equivalent to the 3.5 kb band from Figure 3.4, is indicated, as is the 3.2 kb 28S *rrn*-containing *Eco*RI band. Lane M is the λ *Hin*dIII size marker.

### 3.1.5 HETEROGENEITY IS NOT A GENERAL FEATURE OF THE Lp1 GENOME

It is possible that the length heterogeneity observed in the rDNA IGS of Lp1 is indicative of a high level of genomic variation in general. To address this, a chromosomal karyotype analysis of a selection of Lp1 laboratory cultures and their respective single-spore purified isolates was performed using contour-clamped homogenous electric field (CHEF) pulsed field gel electrophoresis. No differences were observed in the chromosomal banding patterns of any of these cultures within the resolution of this technique (Figure 3.6A).

Secondly the Southern blots shown in Figure 3.2 were stripped and reprobed with an Lp1 *pyr*4-2 clone (λMC11; Collett *et al.*, 1995). Hybridising bands of 3.4 kb, 4.4 kb, 8.5 kb, ~ 20 kb and > 20 kb were observed, consistent with expected results (Figure 3.6B). Weakly hybridising bands were also observed at 3.0 kb and 9.7 kb, and these are believed to result from hybridisation to the second copy of *pyr*4 present in the hybrid, *pyr*4-1 (Collett, 1994). Once again no evidence of length heterogeneity was observed in any of these cultures (the Southern blot used in Figure 3.2A is shown in Figure 3.6B as representative of these results). No evidence of heterogeneity was reported when Lp1 genomic DNA was probed with the *hmg* gene (Dobson, 1997). Thus the length heterogeneity observed with the ribosomal probes does not appear to be a general feature of the Lp1 genome.

### 3.1.6 INTERGENIC SPACER CONTAINS SUB-REPEAT ELEMENTS

To investigate the nature of the length heterogeneity, the 4.1 kb *Sal*I IGS insert from pPN50 was sequenced. I carried out an exonuclease III deletion strategy to sequence across the whole insert in pPN50, as set out in Section 2.15. pPN50 DNA was digested with *Sst*I and *Bam*HI to give a 7.3 kb linear piece of DNA, and then subject to exonuclease treatment, with samples being removed every 70 sec. The progressive digestion of pPN50 by exonuclease III can be seen in Figure 3.7A. These deleted pPN50 clones were circularised and transformed, and a number of clones were selected to try and give 300 bp nested deletions covering the entire 4.1 kb *Sal*I IGS insert. The first group of clones selected to cover the 4.1 kb *Sal*I IGS insert is shown in Figure 3.7B. These were then sequenced.

This resulted in single-stranded sequence covering most of the 4.1 kb *Sal*I IGS clone. There was one gap in the sequence, and this was filled by designing primers nts7 and nts8 to the flanks of the gap. These primers were used to PCR amplify the missing sequence, and the PCR product was directly sequenced to give the complete (barring some repetitive elements - see Section 3.1.6 for details), primarily single-stranded, sequence of the 4.1 kb IGS clone. Sequence was also obtained for the edges of the 5.6 kb *Sal*I coding region insert from pPN49. The sequence of the IGS region and flanks is presented in Figure 3.8. This figure shows the location of all the IGS primers used in this study, as well as other pertinent features that are discussed in subsequent sections. Also shown is a schematic diagram of the organisation of the IGS in Lp1.

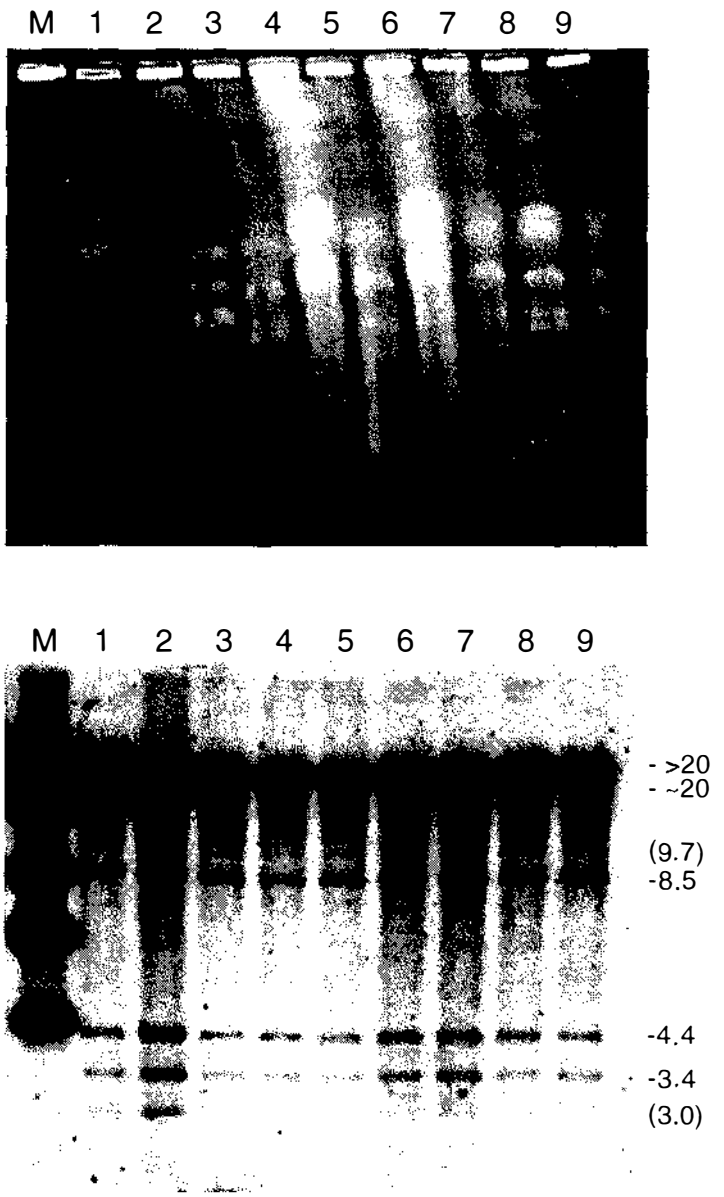



Figure 3.6 Length heterogeneity is not a general feature of the Lp1 genome

A Ethidium bromide stained 0.6% chromosomal grade agarose CHEF gel of various Lp1 isolates. See Section 2.19.4 for the CHEF conditions used. Isolates are: E8 (lane 1), Lp5 (lane 2), Lp1A0 (lane 3), Lp1D6 (lane 4), Lp1F0 (lane 5), Lp1F3 (lane 6), Lp1F6 (lane 7), Lp2G3 (lane 8) and Lp2G6 (lane 9). Lane M is the *Schizosaccharomyces pombe* chromosomal marker.

B The *Sal*I Southern blot from Figure 3.3A was stripped and reprobed with a fragment containing *pyr*4-2 from Lp1 (λMC11; Collett *et al.*, 1995). Isolates are Lp1A0 (lane 1), Lp1A3 (lane 2), Lp1A6 (lane 3), Lp1C0 (lane 4), Lp1C3 (lane 5), Lp1C6 (lane 6), Lp1D0 (lane 7), Lp1D3 (lane 8) and Lp1D6 (lane 9). Lane M is the λ *Hin*dIII size marker. The sizes of the expected fragments are indicated in kb according to Collett (1994). The bands indicated by sizes in brackets correspond to the sizes of bands expected for the *pyr*4-1 copy present in Lp1.

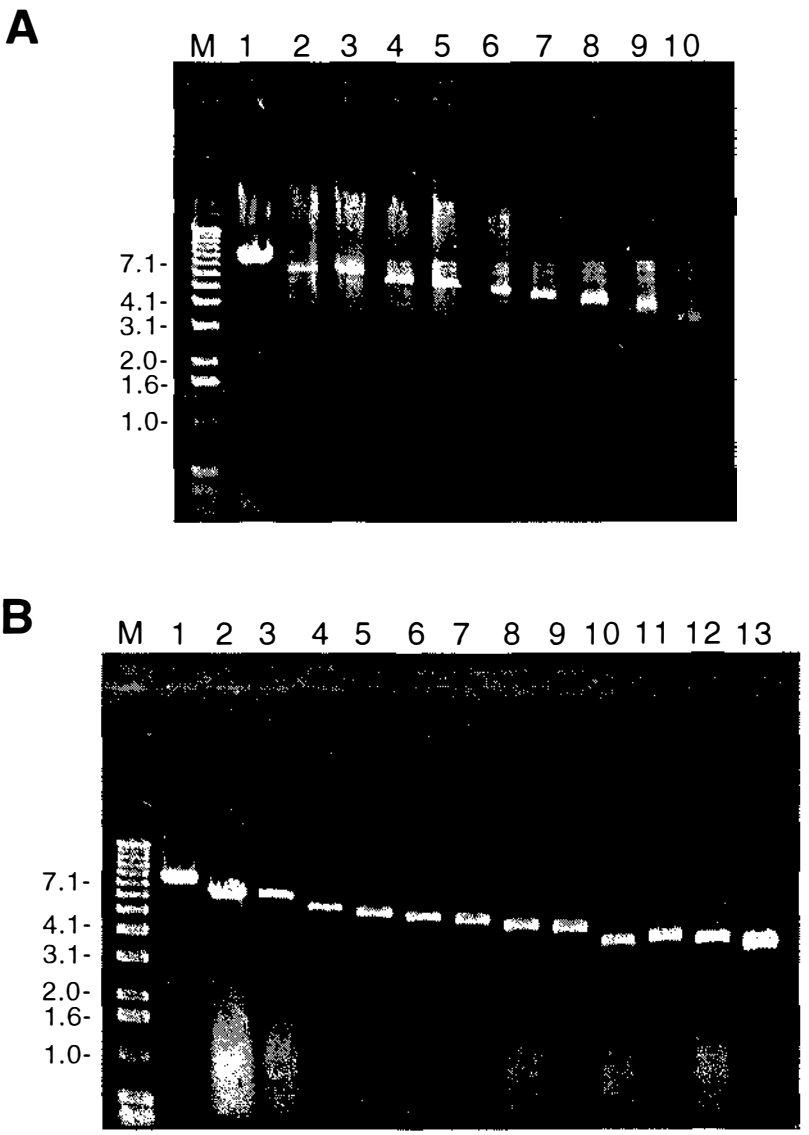
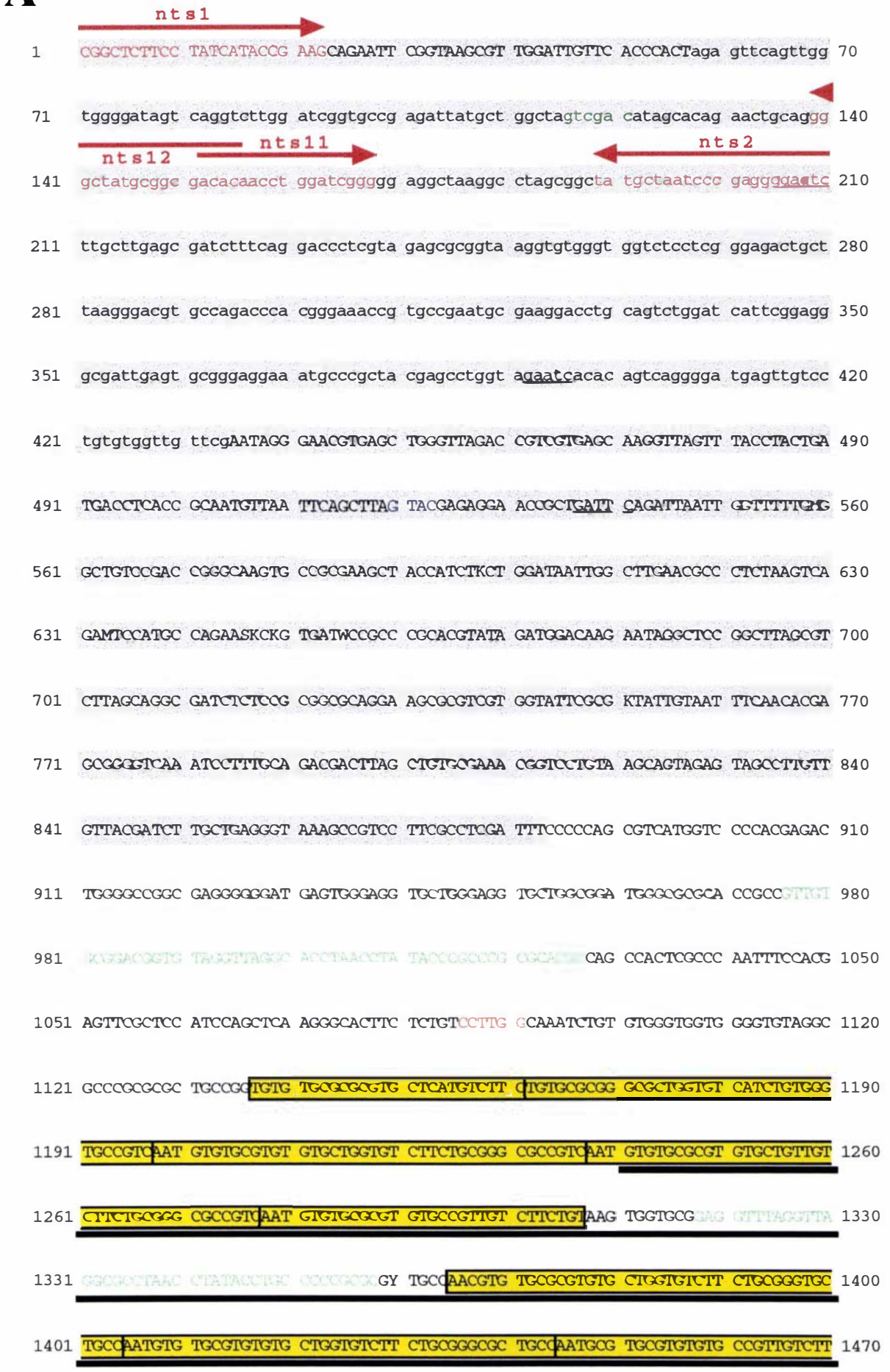


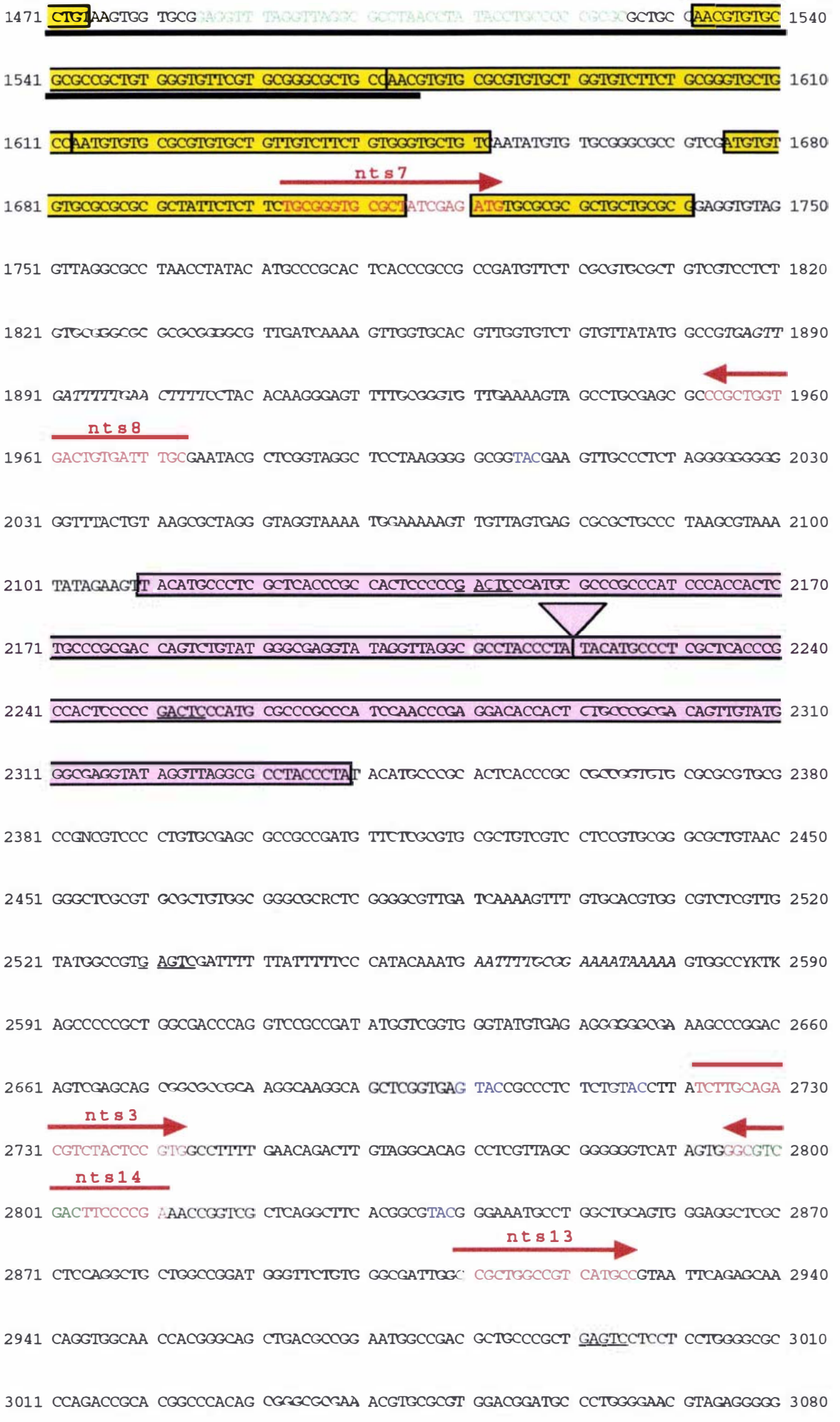

Figure 3.7 Generation of nested deletions of the Lp1 IGS clone by exonuclease III digestion

A Ethidium bromide-stained 0.7% agarose gel showing the time course of exonuclease III digestion of pPN50 which has been linearised with *Sst*I and *Bam*HI. Lane 1 is linearised, intact pPN50. Samples were removed at 70 sec intervals to give a pool of nested deletions from 70 to 630 sec, and these are shown in lanes 2 to 10. Lane M is the 1 kb ladder size marker (Gibco BRL) and the sizes in kb are shown to the left.

B Ethidium bromide-stained 0.7% agarose gel showing a representative group of nested deletion clones from panel A that were ligated, transformed and selected to give coverage of the insert in pPN50, and then the clones were sequenced. The clones have been linearised with *Hin*dIII and differ in size by approximately 300 bp (lanes 2 to 13). Lane 1 is intact pPN50. The samples Lane M is the 1 kb ladder size marker (Gibco BRL) and the sizes in kb are shown to the left.

**A**


nts1
nts12
nts11
nts2
1 CGGCTCTTCC TATCATACCG AAGCAGAATT CGGTAAGCGT TGGATTGTTC ACCCACTaga gttcagttgg 70
71 tggggatagt caggtcttgg atcggtgccg agattatgct ggctagtcga catagcacag aactgcaggg 140
141 gctatgcggc gacacaacct ggatcggggg aggctaaggc ctagcggcta tgctaatccc gagggaactc 210
211 ttgcttgagc gatctttcag gaccctcgta gagcgcggta aggtgtgggt ggtctcctcg ggagactgct 280
281 taagggacgt gccagaccca cgggaaaccg tgccgaatgc gaaggacctg cagtctggat cattcggagg 350
351 gcgattgagt gcgggaggaa atgcccgcta cgagcctggt agaatcacac agtcagggga tgagttgtcc 420
421 tgtgtggttg ttcgAATAGG GAACGTGAGC TGGGTTAGAC CGTCGTGAGC AAGGTTAGTT TACCTACTGA 490
491 TGACCTCACC GCAATGTTAA TTCAGCTTAG TACGAGAGGA ACCGCTGATT CAGATTAATT GGTTTTTGAG 560
561 GCTGTCCGAC CGGGCAAGTG CCGCGAAGCT ACCATCTKCT GGATAATTGG CTTGAACGCC CTCTAAGTCA 630
631 GAMTCCATGC CAGAASKCKG TGATWCCGCC CGCACGTATA GATGGACAAG AATAGGCTCC GGCTTAGCGT 700
701 CTTAGCAGGC GATCTCTCCG CGGCGCAGGA AGCGCGTCGT GGTATTCGCG KTATTGTAAT TTCAACACGA 770
771 GCGGGGTCAA ATCCTTTGCA GACGACTTAG CTGTGCGAAA CGGTCCTGTA AGCAGTAGAG TAGCCTTGTT 840
841 GTTACGATCT TGCTGAGGGT AAAGCCGTCC TTCGCCTCGA TTTCCCCCAG CGTCATGGTC CCCACGAGAC 910
911 TGGGGCCGGC GAGGGGGGAT GAGTGGGAGG TGCTGGGAGG TGCTGGCGGA TGGGCGCGCA CCGCCGTTGT 980
981 XGGACGGTG TAAGTTAGGC ACCTAACCTA TACCGGCCCG CGCAGTC CAG CCACTCGCCC AATTTCCACG 1050
1051 AGTTCGCTCC ATCCAGCTCA AGGGCACTTC TCTGTCCTTG GCAAATCTGT GTGGGTGGTG GGGTGTAGGC 1120
1121 GCCCGCGCGC TGCCGCTGTG TGCGCGCGTG CTCATGTCTT CTGTGCGCGG GCGCTGGTGT CATCTGTGGG 1190
1191 TGCCGTCAAT GTGTGCGTGT GTGCTGGTGT CTTCTGCGGG CGCCGTCAAT GTGTGCGCGT GTGCTGTTGT 1260
1261 CTTCTGCGGG CGCCGTCAAT GTGTGCGCGT GTGCCGTTGT CTTCTGTAAG TGGTGCG GAG GTTTAGGTTA 1330
1331 GGCGGCTAAC CTATACCTGC CCCCGCCX GY TGCCAACGTG TGCGCGTGTG CTGGTGTCTT CTGCGGGTGC 1400
1401 TGCCAATGTG TGCGTGTGTG CTGGTGTCTT CTGCGGGCGC TGCCAATGCG TGCGTGTGTG CCGTTGTCTT 1470


nts7
nts8
nts3
nts14
nts13

3081 AAAGCGGAAT GCCACAGGCG GTCTCCCCGG CCCAAAGCCG GCCGAGGCCA CCCGCGGCTC GCCCGTGTTA 3150

3151 CAGCGGGCGA CCACCAGAAG ACTAACTCTC TCCAGCGCCG CTCCGGCGGC GCTGGCAGGC CACCCCCGGC 3220

nts4

3221 GAACGATAGT TACCTGGTTG ATGTAGTCAT ATGCTTGTCT C

**B**

1 TACATGCCCT CGCTCACCCG CCACTCCCCC GACTCCCATG 40

1 TACATGCCCT CGCTCACCCG CCACTCCCCC GACTCCCATG 40

111-R

41 CGCCCGCCCA TCC. . . . . ..CACCACT CTGCCCGCGA 70

111-L

119-R

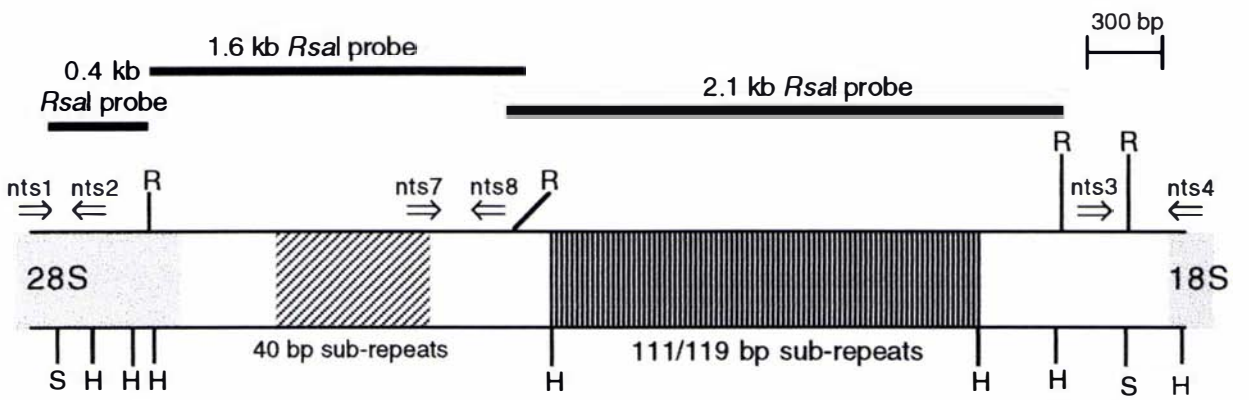
41 CGCCCGCCCA TCCAACCCGA GGACACCACT CTGCCCGCGA 80

119-L

71 CCAGTCTGTA TGGGCGAGGT ATAGGTTAGG CGCCTACCCT A 111

81 C.AGT.TGTA TGGGCGAGGT ATAGGTTAGG CGCCTACCCT A 119

**C**



Analysis of the IGS sequence revealed two sub-repeat classes within the IGS (Figure 3.8). The first, termed the 40 bp sub-repeat class, is a relatively heterogeneous class with a core consensus of 40 bp. The individual repeats of this class are organised in a head-to-tail tandem array, with thirteen either full-length or partial repeats present in the 4.1 kb *Sal*I IGS clone. This sub-repeat class is characterised by alternating pyrimidine-purine residues. The tandem array is not continuous, with a 53 bp unrelated sequence being present at two points within the array. Interestingly, this interrupting sequence is also present once upstream of the 40 bp sub-repeats, and 41 bp of it shows 88% identity to the 111/119 bp sub-repeats discussed below.

The second sub-repeat class is termed the 111-119 bp sub-repeat class. This class is actually composed of two very closely-related sub-repeats, one 111 bp in length and the other 119 bp in length, and these show a high level of identity to each other and to themselves (Figure 3.8B). They are also organised in a head-to-tail tandem array. The sub-repeats of this class are GC-rich, containing on average 65% GC. I was not able to determine the number of these sub-repeats present in the 4.1 kb *Sal*I IGS clone by sequencing, but subsequent analysis demonstrated that it contains 14 sub-repeats (Section 3.1.8). Similarly I was not able to determine the exact sequence of all these 14 repeats, and therefore the consensus sequence of the two sub-repeat variants is presented in Figure 3.8.

### 3.1.7 DIGESTION OF THE 111/119 BP SUB-REPEATS ABOLISHES HETEROGENEITY

Length heterogeneity in the IGS of other species studied results from variation in the number of sub-repeats within the IGS (see Section 4.1 for citations). I suspected that the length heterogeneity observed in the IGS of Lp1 was also the result of IGS sub-repeat copy number variation; a likely candidate being the longer 111/119 bp sub-repeats. To determine whether this sub-repeat class is actually responsible for the observed length heterogeneity, I identified a restriction endonuclease, *Hin*fI, that cleaves these sub-repeats once but leaves the 40 bp sub-repeats intact (Figure 3.8). If copy number variation of the 111/119 bp sub-repeats is responsible for the length heterogeneity, cleaving them with *Hin*fI should abolish the heterogeneity, leaving a high copy-number band the size of the sub-repeats.

To simplify analysis of the results, I identified a restriction endonuclease, *Rsa*I, that cleaves the IGS insert in pPN50 into three fragments of 0.4 kb, 1.5 kb and 2.1 kb suitable for probes, as well as other, smaller fragments (Figure 3.8A). The *Sal*I insert from pPN50 was gel extracted and digested with *Rsa*I, and the three larger fragments were gel extracted for use as probes (Figure 3.8C). Genomic DNA from two representative Lp1 single-spore purified cultures (Lp1A3 and Lp1C3), and from Lp5 and E8 was cleaved with *Hin*fI, separated on an agarose gel and the Southern blot was probed sequentially with the three *Rsa*I sub-fragments (Figure 3.9). None of the three *Rsa*I probes reveal any evidence of the length heterogeneity seen in Figure 3.2. The bands present are all of the sizes predicted from the sequence of the 4.1 kb IGS clone. The probe covering the region that includes the 111-119 bp sub-repeats is the 2.1 kb *Rsa*I probe.

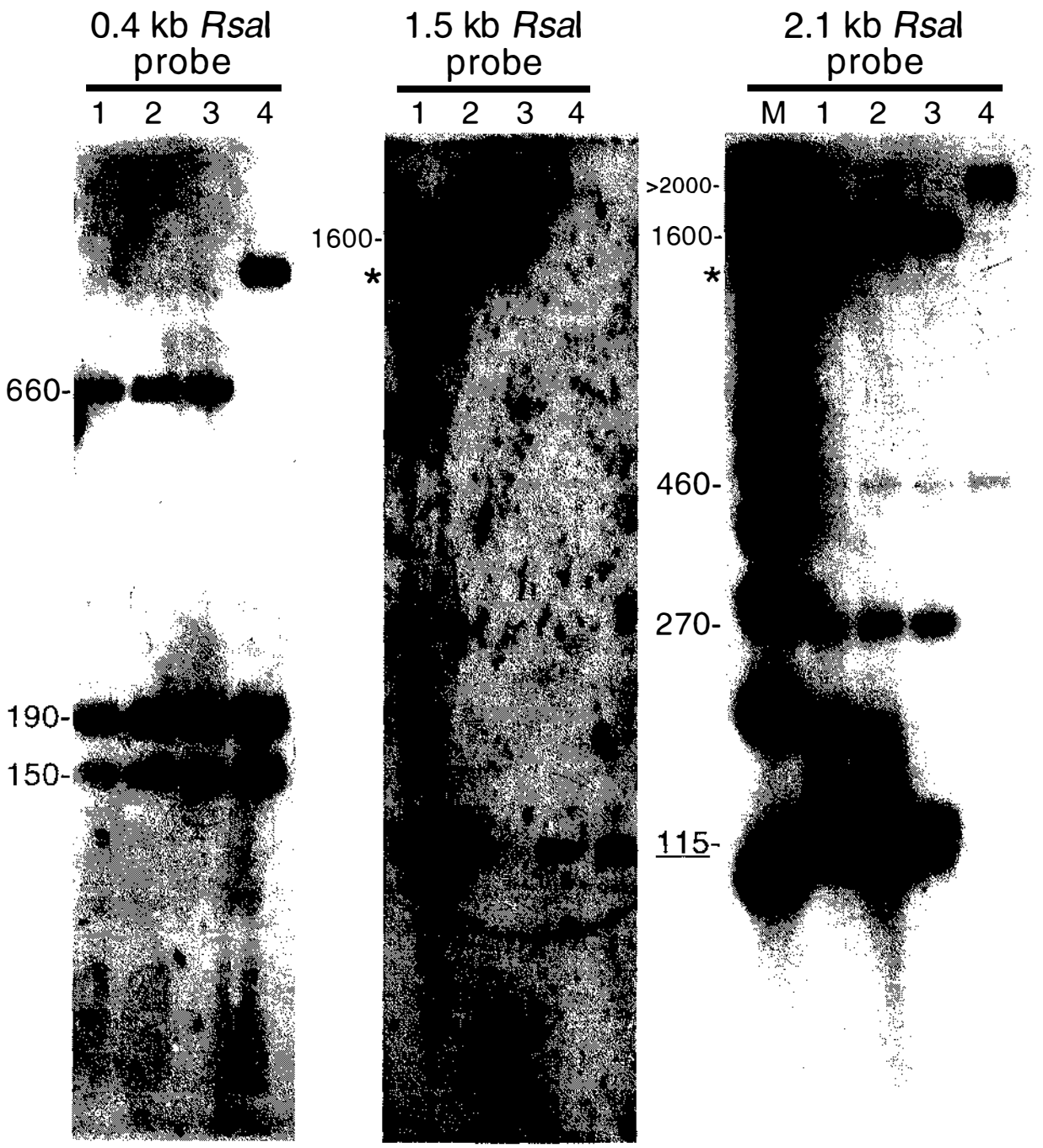



Figure 3.9 Length heterogeneity is the result of copy number variation of the 111/119 bp sub-repeats

1.2 µg of genomic DNA from Lp1A3 (lane 1), Lp1C3 (lane 2), E8 (lane 3) and Lp5 (lane 4) was digested with *Hin*fI and fractionated on a 3% NuSieve agarose gel. The resulting Southern blot was probed with the three *Rsa*I fragment probes from the *Sal*I IGS clone as indicated. Sizes of the hybridising bands in Lp1 are indicated in bp. The 1.27 kb unexpected band is marked with *. Lane M is the 100 bp ladder size marker (Gibco BRL).

There is no evidence of length heterogeneity in this region after *Hin*fI digestion, and as predicted there is a strongly-hybridising band the size of a single sub-repeat (~ 115 bp). In one of the Lp1 laboratory cultures (Lp1A3) an unexpected band of 1.27 kb is found (marked * in Figure 3.9, lane 1). This appears to be a length polymorphism in the spacer that is not the result of the 111/119 bp sub-repeats, and is discussed further below. However the presence of this size polymorphism does not explain the level of the IGS length heterogeneity seen in Figure 3.2.

Abolition of length heterogeneity with *Hin*fI digestion demonstrates that the IGS length heterogeneity is the result of copy number variation of the 111-119 bp IGS sub-repeats.

Probing E8 genomic DNA with the 2.1 kb *Rsa*I fragment produces the same bands as seen in Lp1, except that the intensity of hybridisation of the band around 115 bp is not as strong (Figure 3.9, lane 3). Lp5 produces a somewhat different pattern of bands to Lp1 and E8, and there does not appear to be any hybridisation around 115 bp (Figure 3.9, lane 4). Thus the IGS structure in Lp5 appears to differ from that found in Lp1 and E8 (see Section 3.2.1).

### 3.1.8 ARRANGEMENT OF THE 111/119 BP SUB-REPEATS IN THE IGS

Information on the numbers and distribution of repeats within an array can give insights into the processes that are shaping the array, in particular the mechanism (s) responsible for their homogenisation. I employed a technique called multi-variant repeat PCR (MVR-PCR; Jeffreys *et al.*, 1991) to determine the order of the two types of sub-repeat in the 111/119 bp sub-repeat array (Dover *et al.*, 1993). The strategy I used to determine the order of the sub-repeat array is shown schematically in the upper part of Figure 3.10 and is as follows (see also Section 2.18). Primers are designed to each of the sub-repeat types in the array (111 and 119 sub-repeat-specific primers) so they specifically amplify that sub-repeat type. At the 5' end of each primer is a sequence complementary to a random sequence primer that does not anneal to anything in the genome (the TAG primer). A single cycle of PCR amplification is carried out using the sub-repeat-specific primers (in separate reactions) at low concentrations, and a primer outside the sub-repeat array (the anchor primer; see below) at a standard concentration. Following this single cycle the TAG primer is added at a standard concentration, and PCR amplification is performed for another 25 cycles. In the initial PCR cycle the sub-repeat-specific primers anneal to their respective sub-repeats and produce a series of amplification products corresponding to each sub-repeat of that type in the array. These products now have a TAG priming site courtesy of the sub-repeat-specific primer (Figure 3.10). The subsequent rounds of PCR are biased towards amplification from the TAG and anchor primers to give stable amplification of the products (from all the sub-repeat units in the array) and allowing the order of the two sub-repeat types to be determined from the sizes of the PCR products. PCR without the TAG primer step is expected to give collapse of the amplification products down to a single product of the size expected for the first sub-repeat in the array. The PCR products are run on an agarose gel, and the order of sub-repeats is determined by reading the order of ethidium bromide-stained bands up the gel.

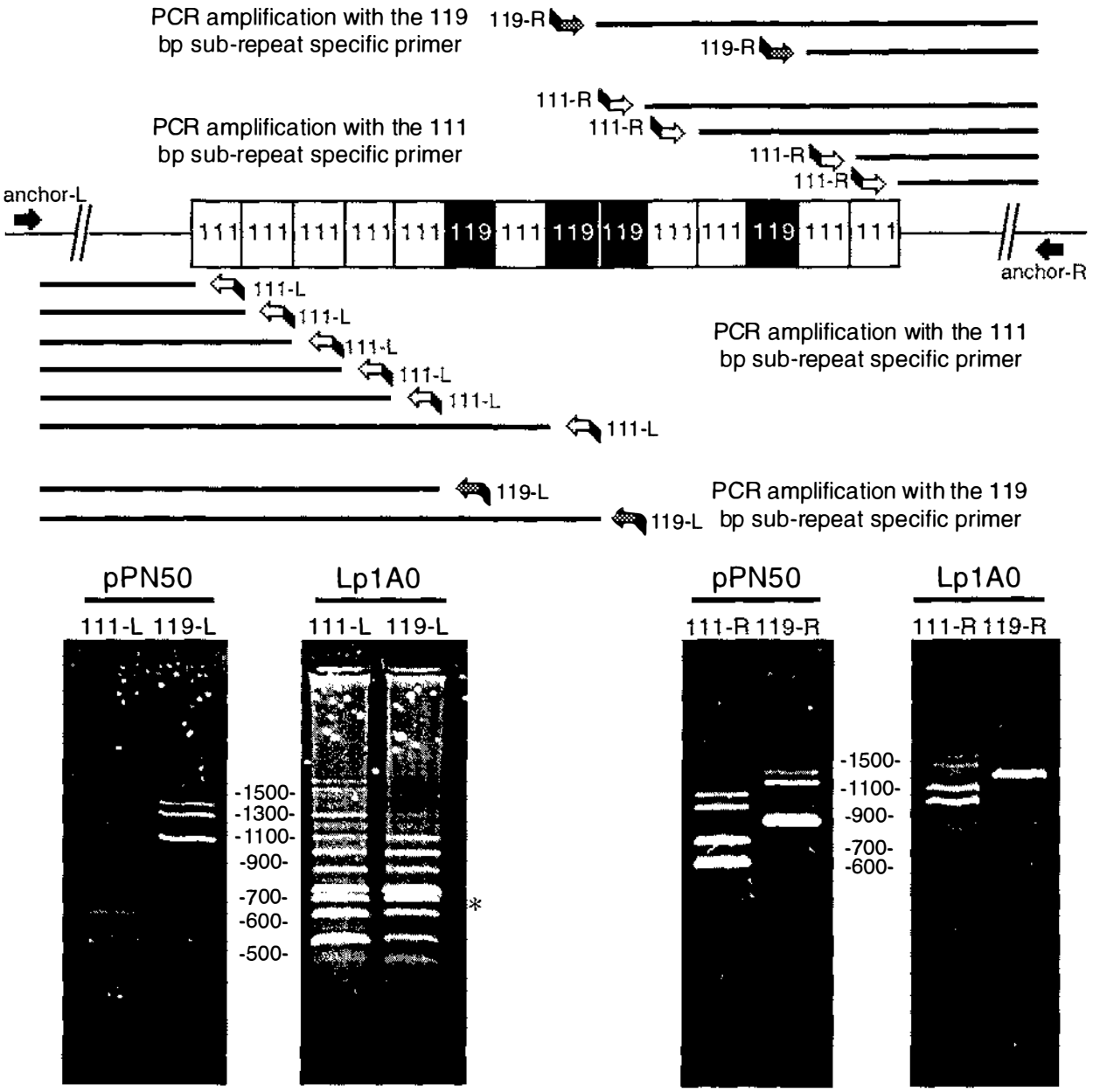



Figure 3.10 Multi-variant repeat PCR mapping of the 111/119 bp sub-repeats in the Lp1 rDNA IGS

The upper part of the figure shows the multi-variant repeat PCR analysis scheme. The 111 bp sub-repeat variants are boxed in white and the 119 bp sub-repeat variants are boxed in black. The first round of PCR amplification is shown from each end of the sub-repeat array. The 111 bp (open arrows) and 119 bp (stippled arrows) sub-repeat specific primers anneal to their respective sub-repeat variant. In combination with the appropriate anchor primer they amplify the intervening region, with the ladder of products shown as thick lines. These products are then stably amplified using the TAG primer (filled tail on sub-repeat specific primers) and anchor primer. Diagram not drawn to scale. The lower part of the figure shows ethidium bromide stained 1.5% agarose gels with the amplification products from pPN50 or Lp1A0 genomic DNA as indicated. Reading the ladder of bands up from each pair of sub-repeat specific reactions gives the sub-repeat order, and the results from the two halves of the array can be overlapped. Sizes of the 100 bp ladder size marker (Gibco BRL) are indicated in bp. The unexpected band seen with genomic DNA is marked by *.

I found that the High Fidelity™ *Taq* DNA polymerase (Boehringer-Mannheim) used in the PCR reactions had difficulty in traversing these sub-repeats. Hence two MVR-PCR reactions, initiating from each end of the sub-repeat array, were needed to determine the order of the entire array in the clone, and therefore two sub-repeat-specific primers were required for each sub-repeat (111-L and 111-R, and 119-L and 119-R). I carried out MVR-PCR analysis using both the Lp1 clone (pPN50) and genomic DNA as templates. The anchor primers, located to the left and right of the sub-repeat array, were nts7 for anchor-L, and nts4 and the M13-forward primer for anchor-R in genomic and clone DNA respectively. Combining the results from each end of the array and finding the overlap gives the complete order of sub-repeats in the clone.

The results are shown in the lower part of Figure 3.10. The order of the sub-repeats in pPN50 could be unambiguously determined, despite the lack of specificity of the 111-L primer (see below). The order of sub-repeats found in the clone is that indicated in the upper part of Figure 3.10. No apparent pattern of organisation for these two sub-repeat variants is evident. The 111 bp sub-repeat is the predominant type.

The 111-L sub-repeat-specific primer is not specific for the 111 bp sub-repeats, but amplifies both the 111 bp and 119 bp sub-repeats equally well (Figure 3.10). This is likely to be a consequence of the primer sequence, as the only difference in sequence between the two variant sub-repeats recognised by this primer is the two 3'-most bases (Table 2.2 and Figure 3.8B). This difference does not appear to be sufficient to distinguish the two sub-repeats with the PCR conditions used. Raising the annealing temperature abolished amplification (data not shown), presumably because the anchor-L primer (nts7) failed to anneal. This lack of specificity does not prevent the ordering of the sub-repeats because the specificty of the 119-L sub-repeat-specific primer clearly shows the order.

The results for the MVR-PCR with Lp1A0 genomic DNA are also shown in the gels in Figure 3.10. The results for genomic DNA give some idea about the conservation or lack thereof of sub-repeat order in the IGS within the rDNA cluster. If the order of sub-repeats from genomic DNA can be clearly ordered, then most of the rDNA units have the same order of sub-repeats. Similarly a lack of order of sub-repeats in genomic DNA indicates that different rDNA units have different orders of the sub-repeats. The bands from the right hand side of the sub-repeat array can be ordered for about six sub-repeats, and this order is the same as in the clone. The specificity of banding is not as clear as in the clone, indicating some heterogeneity of sub-repeat order exists amongst the population of IGS, but nevertheless an unambiguous order can be determined. Conversely, the results for the left hand side of the sub-repeat array do not resemble the clone, and there does not appear to be any clear ordering of the sub-repeats at this edge of the array. Once again the lack of specificity of 111-L is not likely to confound the results as gaps in the lane with 119-L would be expected. Instead, the 119-L primer anneals to many more sub-repeats when genomic DNA is used as the template. This indicates that the population of IGS has considerable variation in the order of sub-repeats at the left-hand end of the sub-repeat array.

The two sides of the array do not behave in the same way with respect to maintenance of sub-repeat order. Another feature not found with the clone is the presence of an extra band in the ladder of sub-repeats on the left-hand side of the array (marked with * in Figure 3.10). It is difficult to know the cause of this extra band, as it is not present in the clone despite its apparent ubiquity in genomic DNA. It may result from a rearrangement in one of the sub-repeats, such that the priming site has been duplicated so one sub-repeat has two priming sites.

### 3.1.9 IGS LENGTH HETEROGENEITY IN OTHER *EPICHLOË* ENDOPHYTE HYBRIDS

The extraordinary level of IGS length variation is only observed in the hybrid, Lp1, not in the two putative progenitors. This suggests that the generation of the length heterogeneity is somehow associated with the hybridisation event. To determine whether a dramatic increase in IGS length heterogeneity is a common occurrence in *Epichloë* endophyte hybrids, genomic DNA from the putative endophyte hybrids FaTG-3 (isolate Tf16), *Neotyphodium uncinatum* (isolate Fp4), *N. coenophialum* (isolates Tf27 and Tf28) and FaTG-2 (isolates Tf13 and Tf15), as well as from the atypical *N. lolii* LpTG-1 isolate, Lp9 (Moon *et al.*, 1999), was digested with *Sal*I, separated on an agarose gel and the Southern blot was probed with the 4.1 kb *Sal*I IGS probe (Figure 3.11). Lp1D0 is shown alongside for comparison. Hybridisation to these other endophyte isolates was very weak, presumably due to high levels of interspecific divergence in the IGS sequence. However, signals were obtained. In no case was the extraordinary level of IGS length heterogeneity in Lp1 observed for any of these endophyte hybrids. In several cases there were multiple hybridising bands and probing with a homologous probe might reveal some degree of length heterogeneity. However it does not appear that the degree of length heterogeneity seen in Lp1 is typical of *Epichloë* endophyte hybrids in general. Further work is required to assess the level of IGS length variation present in these hybrids.

### 3.1.10 SPREAD OF A DELETION IN THE IGS

I wanted to characterise the polymorphism observed in the *Hin*fI genomic digest (Figure 3.9, marked with *). This appeared to be a polymorphic derivative of the 1.60 kb *Hin*fI band which, according to the Lp1 IGS sequence, is located on the 28S *rrn* side of the 111/119 bp sub-repeat array. PCR amplification with nts1 and nts8 using the same Lp1 genomic DNA samples as those used in Figure 3.9 showed that the polymorphism was a deletion (Figure 3.12A). This deletion was not manifested in a PCR amplification with nts7 and nts8, narrowing down the region of the deletion. To obtain a DNA fragment that could be fully sequenced, the nts1 - nts8 PCR product was digested with a number of restriction endonucleases. One of these, *Bsa*JI, produced fragments covering the region of the deletion that were of a suitable size to sequence. These *Bsa*JI fragments from the deleted (~ 500 bp) and non-deleted (~ 800 bp) samples were then sub-cloned into a pGEM-T vector and sequenced using the M-13 forward and reverse primers. The sequences were aligned to determine the exact nature of the deletion (Figure 3.12B). The deletion is 335 bp in length, and this corresponds to a deletion of six of the 40 bp sub-repeats and

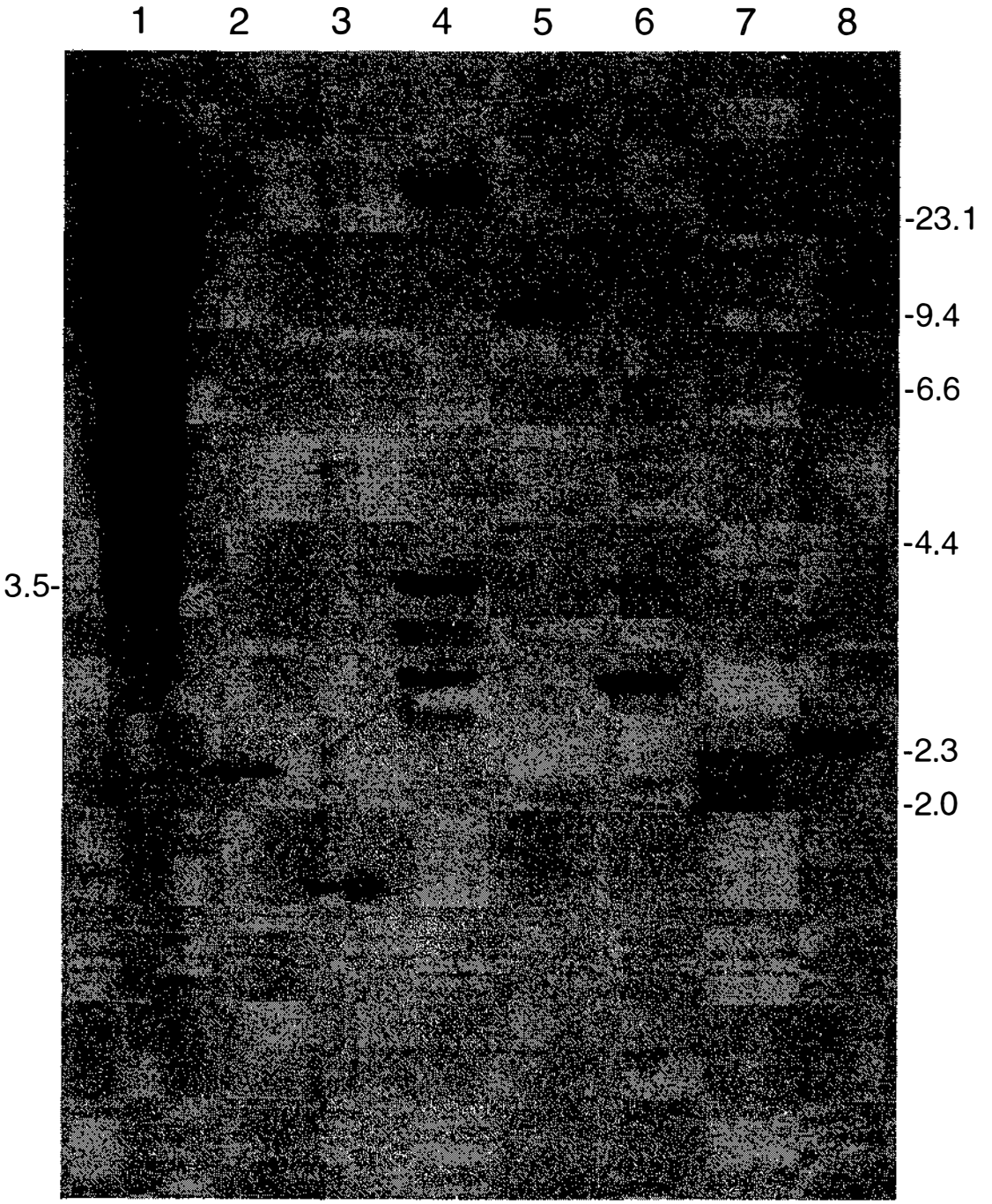



Figure 3.11 IGS length heterogeneity is not a general feature of *Epichloë* endophyte hybrids

1.2 µg of genomic DNA from Lp1D0 (lane 1), *Neotyphodium lolii* LpTG-1 Lp9 (lane 2), FaTG-3 Tf16 (lane 3), *N. uncinatum* Fp4 (lane 4), *N. coenophilum* Tf27 (lane 5), FaTG-2 Tf15 (lane 6), FaTG-2 Tf13 (lane 7) and *N. coenophilum* Tf28 (lane 8) was digested with *Sal*I and fractionated on a 0.7% agarose gel. The resulting Southern blot was probed with the 4.1 kb *Sal*I IGS probe. The 3.5 kb IGS length from Lp1D0 is indicated to the left. The sizes of the λ *Hin*dIII size marker are indicated in kb on the right.

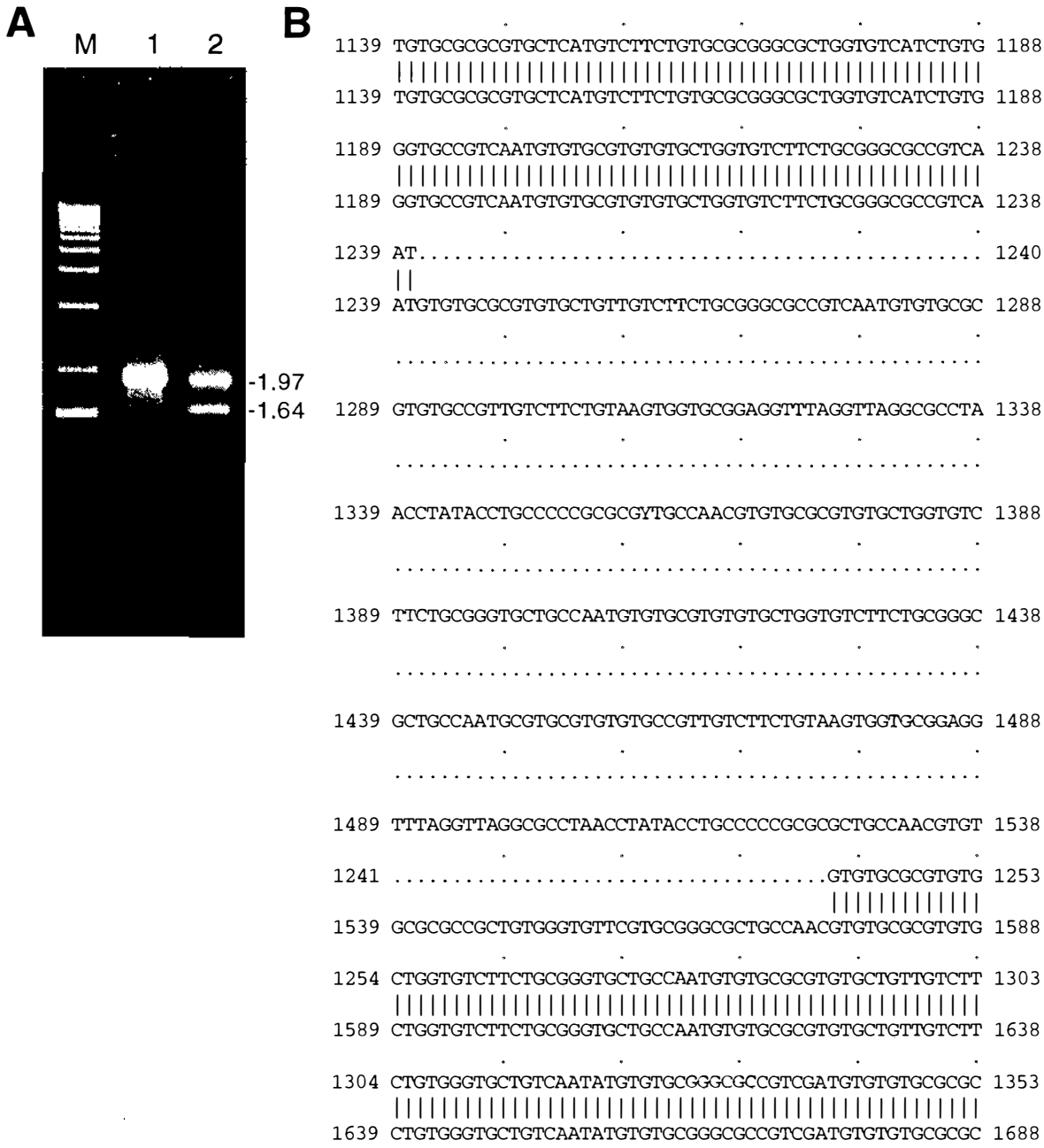



Figure 3.12 Characterisation of the deletion in the Lp1 IGS

A Ethidium bromide stained 0.7% agarose gel showing a deletion of 335 bp in some copies of the IGS in Lp1. PCR was performed on genomic DNA from Lp1D0 (not displaying the deletion; lane 1) and Lp1A3b (displaying a relatively high proportion of the deletion; lane 2) using primers nts1 and nts8. The 1.64 kb deleted and 1.97 kb non-deleted products are indicated to the right of the gel. Lane M is the 1 kb ladder size marker (Gibco BRL).

B Alignment of the *Bsa*JI sub-clones of the deleted (top line) and non-deleted (bottom line) nts1 - nts8 PCR products (shown in panel A). Vertical bars represent matching bases, and dots represent gaps in the alignment. Numbering of the undeleted sequence is in reference to the Lp1 ribosomal sequence presented in Figure 3.8A.

two of the 53 bp interrupting sequences (refer to Figure 3.8A).

The hybridisation signals shown in Figure 3.9 indicate that this deletion derivative is entirely lacking from Lp1C3, but is present in a relatively high proportion in Lp1A3. To examine the origin and possible spread of this deletion derivative, genomic DNA from a number of Lp1 cultures was cleaved with *Hin*fI, separated on agarose gels and the Southern blots were probed with the 4.1 kb *Sal*I IGS probe (Figure 3.13). The results show that the deletion derivative was not present in the original Lp1O isolate from which all the laboratory cultures have been derived (Figure 3.13, lane 1). This indicates the deletion event occurred after isolation of Lp1 from the plant, and it has spread to quite high levels in some of the cultures since. The results also show that the proportion of deletion derivative to full-length sequence varies between cultures. Most interestingly, occasionally the proportion of deletion-containing units changes quite dramatically through one round of single-sporing (e.g. in Figure 3.13, A3 in lane 3 to A3b in lane 6 versus A3d in lane 7 shows a rapid increase in the proportion of non-deleted rDNA).

## 3.2 GROUP I INTRON IN THE LP1 28S *rrn* GENE

### 3.2.1 DISCOVERY OF AN INSERTION ELEMENT IN THE LP1 28S *rrn* GENE

The *in vivo Sal*I junction between the inserts in pPN49 and pPN50 covering the 3' end of the 28S *rrn* gene was spanned using primers designed from the sequence obtained (nts1 and nts2; Table 2.2). PCR amplification of Lp1 genomic DNA using this primer combination produced a product of the expected size, 210 bp, and sequencing confirmed this product contains a *Sal*I site at the appropriate location (Figure 3.8A). The junction between the inserts in pPN49 and pPN50 covering the 5' end of the 18S *rrn* gene (in pPN49) and the 3' end of the IGS (in pPN50) was spanned using a primer that is the reverse complement of ns1 (White *et al.*, 1990; referred to as nts4) and a primer designed from the sequence of the 4.1 kb *Sal*I IGS clone (nts3; Table 2.2). PCR amplification using this primer combination with Lp1 genomic DNA produced a product of the expected size, 479 bp, and sequencing confirmed this product contains a *Sal*I site at the appropriate location (Figure 3.8A). This is further confirmation that the 4.1 kb *Sal*I IGS and 5.6 kb *Sal*I coding region fragments are linked *in vivo*. The sequence of these PCR products was the same as that found in the inserts in pPN49 and pPN50.

Analysis of the regions flanking the IGS revealed an insertion element in the 3' end of the 28S *rrn* gene. There is a 376 bp stretch of sequence that interrupts the homology between the Lp1 28S *rrn* sequence and that from other organisms. An alignment of the 3' region of the 28S *rrn* gene from Lp1, *Saccharomyces cerevisiae*, *Candida albicans*, *Schizosaccharomyces pombe* and rice (*Oryza sativa*) covering this insertion is shown in Figure 3.14. Interestingly, the insertion sequence contains the *Sal*I site that has been exploited in the characterisation of the IGS length variation in this study. As only a single band of the size predicted by the rDNA mapping (Figure

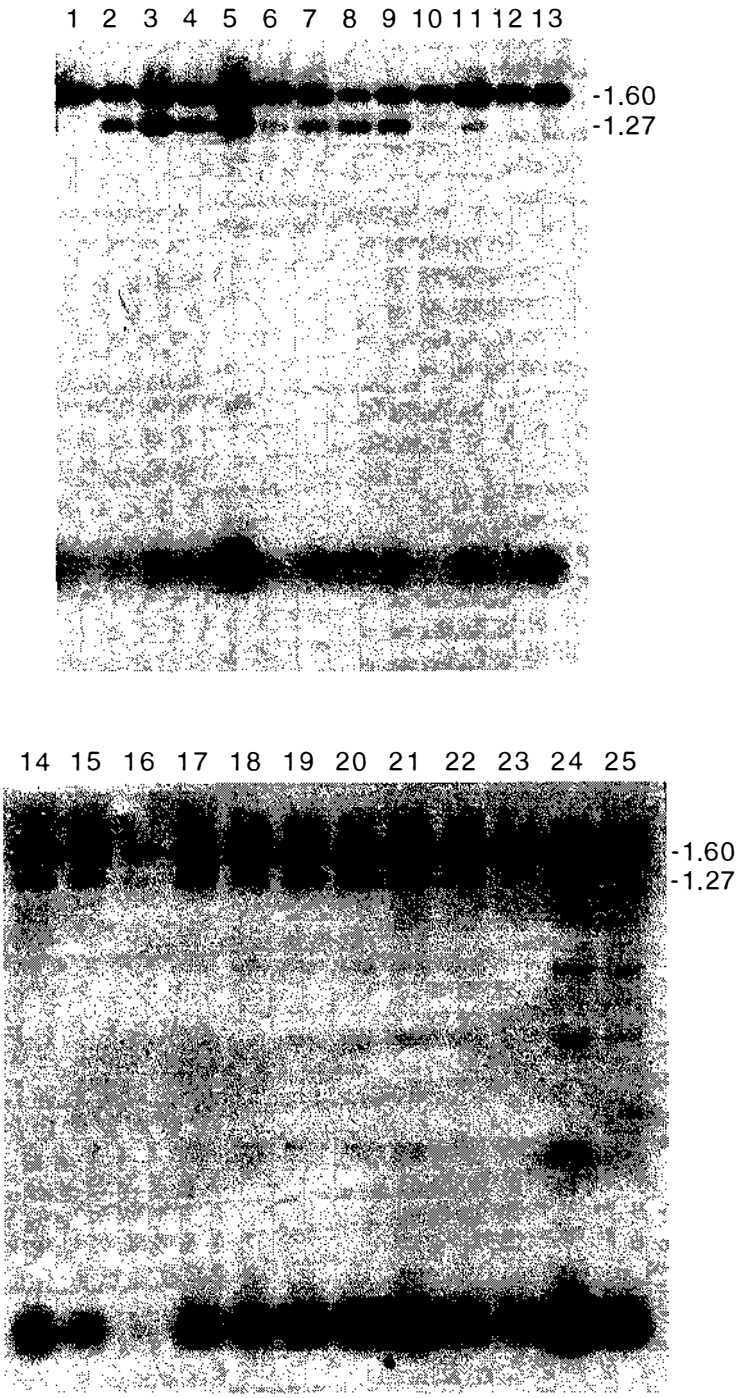



Figure 3.13 Progression of the IGS deletion through rounds of single-sporing

0.8 μg of genomic DNA from Lp1O, Lp1A0, Lp1A3, Lp1A6, Lp1A0 (freshly extracted), Lp1A3b, Lp1A3d, Lp1A6b, Lp1A6d, Lp1C0, Lp1F0, Lp2G0, Lp1D0, Lp1C0, Lp1C3, Lp1C6, Lp1C3b, Lp1C3d, Lp1C6b, Lp1C6d, Lp1F0, Lp1F3, Lp1F6, Lp1A0 and Lp1D0 (lanes 1 to 25 respectively) was digested with *Hin*fI and fractionated on 3% NuSieve agarose gels. The resulting Southern blots were probed with the 4.1 kb *Sal*I IGS probe. Non-deleted (1.60 kb) and deleted (1.27 kb) IGS variant bands are indicated. See Figure 3.10 for explanation of other hybridising bands.

```
                1                                                          60
C.albicans 28S  CGGCTCTTCC TATCATACCG AAGCAGAATT CGGTAAGCGT TGGATTGTTC ACCCACT...
S.cerevisiae 28S CGGCTCTTCC TATCATACCG AAGCAGAATT CGGTAAGCGT TGGATTGTTC ACCCACT...
S.pombe 28S     CGGCTCTTCC TATCATACCG AAGCAGAATT CGGTAAGCGT TGGATTGTTC ACCCACT...
Rice 28S        CGGCTCTTCC TATCATTGTG AAGCAGAATT CACCAAGTGT TGGATTGTTC ACCCACC...
Lp1 28S         CGGCTCTTCC TATCATACCG AAGCAGAATT CGGTAAGCGT TGGATTGTTC ACCCACTAGA

               61                                                         120
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         GTTCAGTTGG TGGGGATAGT CAGGTCTTGG ATCGGTGCCG AGATTATGCT GGCTAGTCGA

               121                                                        180
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         CATAGCACAG AACTGCAGGG GCTATGCGGC GACACAACCT GGATCGGGGG AGGCTAAGGC

               181                                                        240
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         CTAGCGGCTA TGCTAATCCC GAGGGGAGTC TTGCTTGAGC GATCTTTCAG GACCCTCGTA

               241                                                        300
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         GAGCGCGGTA AGGTGTGGGT GGTCTCCTCG GGAGACTGCT TAAGGGACGT GCCAGACCCA

               301                                                        360
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         CGGGAAACCG TGCCGAATGC GAAGGACCTG CAGTCTGGAT CATTCGGAGG GCGATTGAGT

               361                                                        420
C.albicans 28S  .......... .......... .......... .......... .......... ..........
S.cerevisiae 28S ......... .......... .......... .......... .......... ..........
S.pombe 28S     .......... .......... .......... .......... .......... ..........
Rice 28S        .......... .......... .......... .......... .......... ..........
Lp1 28S         GCGGGAGGAA ATGCCCGCTA CGAGCCTGGT AGAATCACAC AGTCAGGGGA TGAGTTGTCC

               421                                                        480
C.albicans 28S  .......... ....AATAGG GAACGTGAGC TGGGTTTAGA CCGTCGTGAG ACAGGTTAGT
S.cerevisiae 28S ......... ....AATAGG GAACATGAGC TGGGTTTAGA CCGTCGTGAG ACAGGTTAGT
S.pombe 28S     .......... ....AATAGG GAACGTGAGC TGGGTTTAGA CCGTCGTGAG ACAGGTTAGT
Rice 28S        .......... ....AATAGG GAACGTGAGC TGGGTTTAGA CCGTCGTGAG ACAGGTTAGT
Lp1 28S         TGTGTGGTTG TTCGAATAGG GAACGTGAGC TGGG.TTAGA CCGTCGTGAG CAAGGTTAGT
```

## Figure 3.14 Insertion element in the Lp1 28S *rrn* gene

Alignment of the 3' end of the 28S *rrn* genes from Lp1, *Candida albicans* (Accession number X70659), *S. cerevisiae* (Accession number J01355), *S. pombe* (Accession number Z19136) and *Oryza sativa* (Accession number M11585) showing the 376 bp insertion in the Lp1 28S *rrn* gene. Dots represent gaps in the sequence alignment. Numbering refers to the Lp1 ribosomal sequence presented in Figure 3.8A. The *SalI* site is underlined.

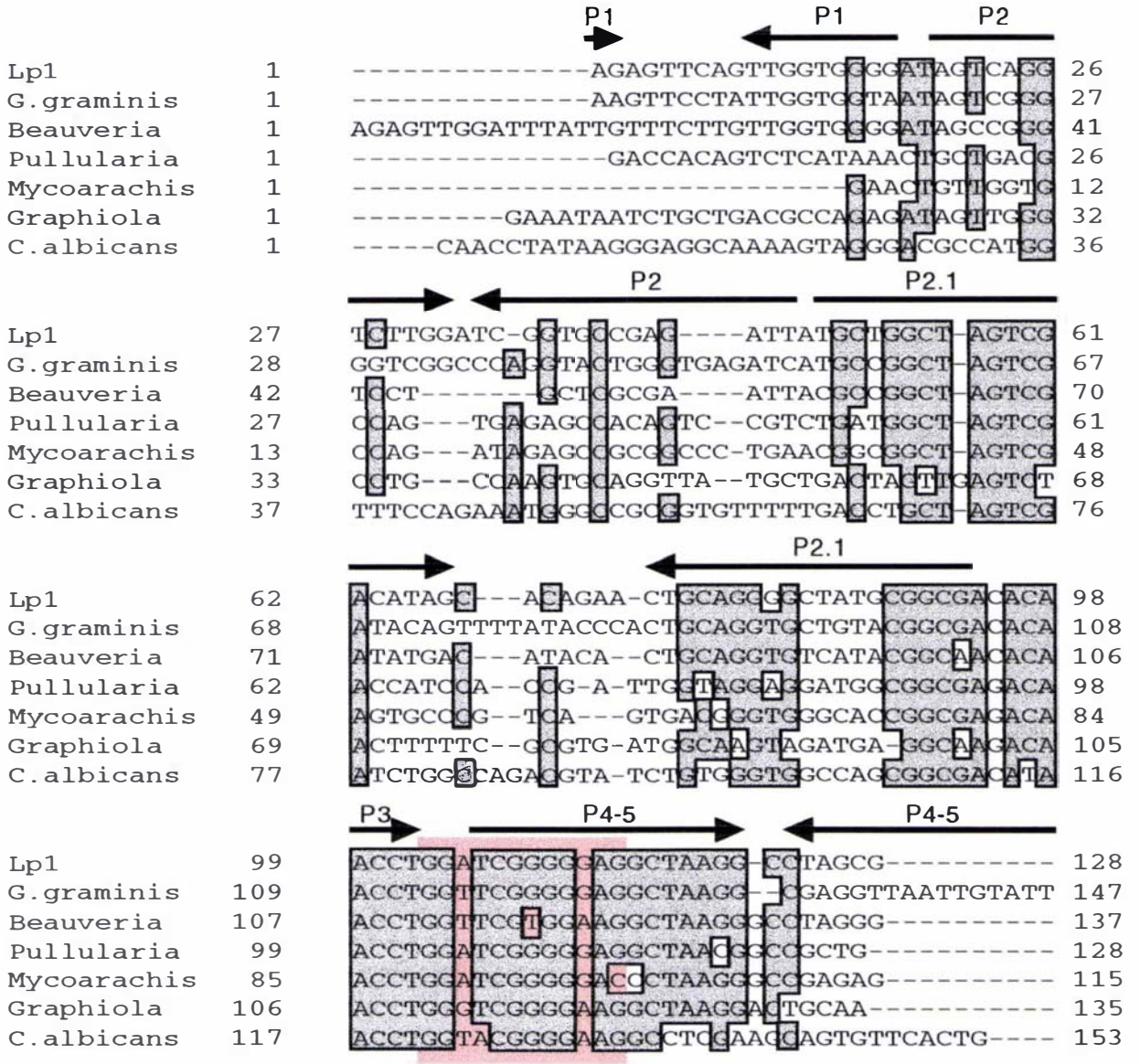
1.1) is seen in *Sal*I digests probed with the 5.6 kb *Sal*I coding region probe (Figures 3.3 and 3.5), this insertion sequence must be present in all copies of the 28S *rrn* gene in Lp1, Lp5 and E8.

### 3.2.2 THE INSERTION ELEMENT IS A GROUP I INTRON

A search of the Genbank, Brookhaven and SWISSPROT databases using the BLAST programme (Altschul *et al.*, 1997) with the sequence of this insertion element revealed similarity to group I introns found in the *rrn* genes (both 18S and 28S *rrn* genes) of other fungi. Of the four group I introns from the database that are most similar to the Lp1 intron, one is located in the 28S *rrn* gene at the same site as the Lp1 intron (*Gaeumannomyces graminis* intron T; Tan, 1997), and the other three are all located in the 18S *rrn* gene (*Pullularia prototropha* intron, *Mycoarachis inversa* intron and *Graphiola phoenicis* intron [Sjamsuridzal and Sugiyama 1998]). I also obtained the sequence of a group I intron from *Beauveria brongniartii* that is inserted in the same location in the 28S *rrn* gene and which shows similarity to the intron from Lp1 (C. Neuvéglise, unpublished results; Neuvéglise *et al.*, 1997). Figure 3.15 shows an alignment of these five group I introns with the intron from Lp1, as well as the other most similar 28S *rrn* group I intron (*Candida albicans* intron; Mercure *et al.*, 1993) as determined by similarity in the P, Q, R and S regions (Tan, 1997). The insertion sites of these introns are given in the legend to Figure 3.15.

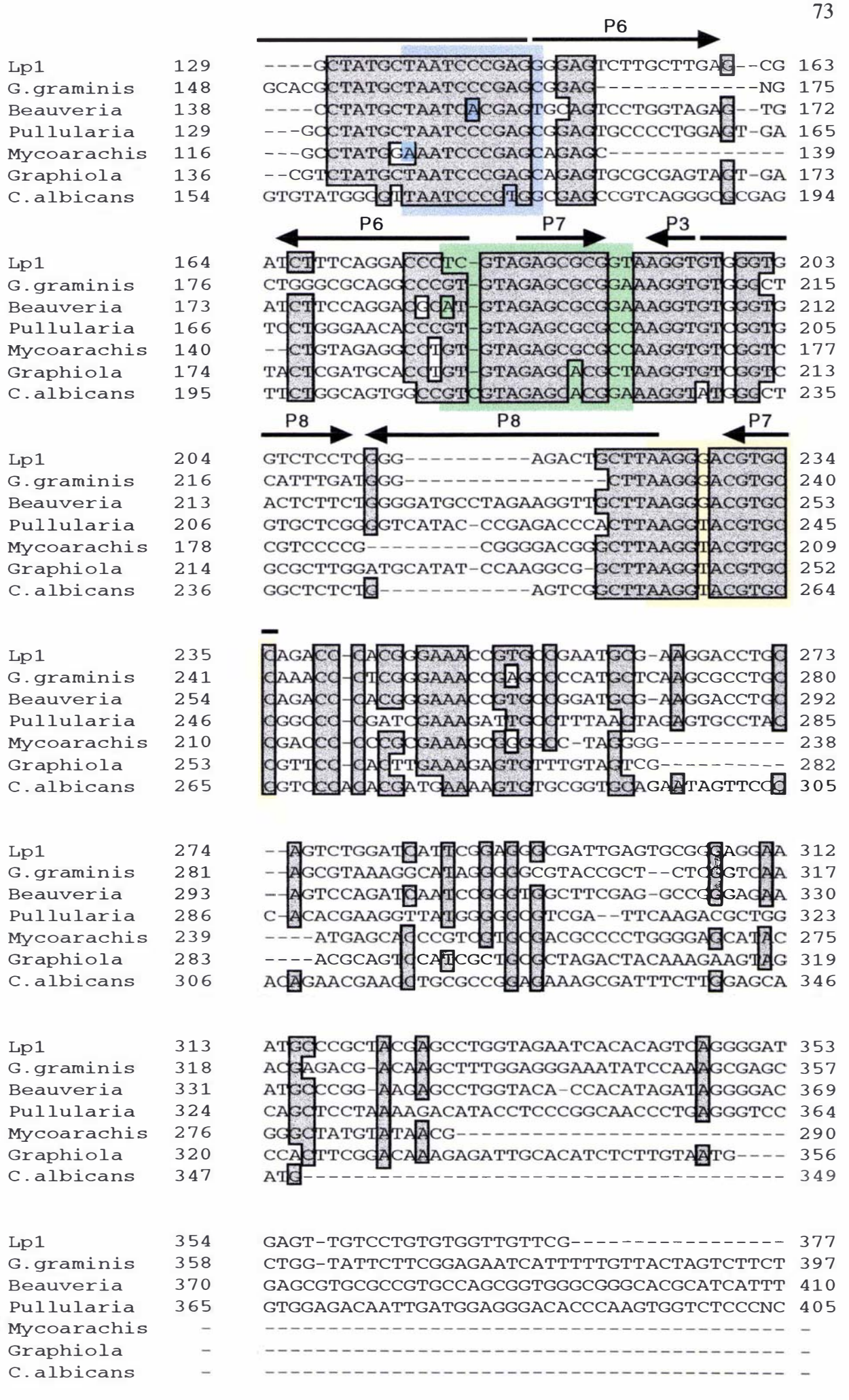
### 3.2.3 STRUCTURE OF THE LP1 GROUP I INTRON

The alignment in Figure 3.15 shows that the similarity between these group I introns lies in restricted blocks. Some of these blocks correspond to the highly-conserved P, Q, R and S elements characteristic of group I introns that together form the catalytic core (Cech, 1988). Using these conserved pairing elements and the common group I intron secondary structure as determined by comparative analyses (Cech, 1990), I determined the putative pairing elements P2 to P8 in the Lp1 group I intron, and these are indicated on Figure 3.15. I did not attempt to determine the P9 pairing element (s), as the sequence conservation was not sufficient to make any firm predictions. These deduced pairing elements fit very well with the common group I intron secondary structure. The 3'-most base of the intron is a G, a feature universally found in group I introns (Cech, 1990). I also identified the putative internal guide sequence consisting of P1 and P10 pairing elements, and these are shown in Figure 3.16. These deduced pairing elements are involved in directing both the 3' and 5' intron splice sites (Davies *et al.*, 1982). The internal guide sequence has the conserved U-G pair at the 5' exon junction (Cech, 1988; Figure 3.16). However, the intron components of the P1 pairing element (responsible for 5' splicing) and the P10 pairing element (responsible for 3' splicing) are located very close together, and it seems unlikely that they would be able to pair simultaneously as shown in Figure 3.16 because of conformational problems.




## Figure 3.15 Group I intron alignment

Alignment of the Lp1 group I intron with introns from *Gaeumannomyces graminis* (Accession number U17161), *Beauveria brongniartii* (C. Neuvéglise, personal communication), *Pullularia prototropha* (Accession number X91899), *Mycoarachis inversa* (Accession number AB012953), *Graphiola phoenicis* (Accession number D63930) and *Candida albicans* (Accession number X74272). Insertion sites (relative to *E. coli rrn* genes) are position 2563 in the the LSU *rrn* gene (Lp1, *G. graminis* and *B. brongniartii*), position 516 in the SSU *rrn* gene (*M. inversa* and *G. phoenicis*) and position 1925 in the LSU *rrn* gene (*C. albicans*). The insertion site of *P. prototropha* is not known. The putative pairing elements P2 through to P8 are indicated by arrows above the alignment. These only refer to the Lp1 sequence. P9 pairing elements are not included because of uncertainty in their pairing, and P1 and P10 pairing elements are shown in Figure 3.16. The P, Q, R and S regions that form the catalytic core of the intron are shown with a red, blue, green and yellow background respectively. Nucleotides that are conserved in 5 of the 8 introns are blocked in gray. Gaps are indicated by dashes. The ends of the introns from *G. graminis*, *B. brongniartii* and *P. prototropha* are not shown in the alignment. For details on how these introns were selected, see text.


P6
P6
P7
P3
P8
P8
P7
Lp1 129 ----GCTATGCTAATCCCGAGGGGAGTCTTGCTTGAG--CG 163
G.graminis 148 GCACGCTATGCTAATCCCGAGCGGAG----------NG 175
Beauveria 138 ----CCTATGCTAATCACGAGTGCAGTCCTGGTAGAG--TG 172
Pullularia 129 ---GCCTATGCTAATCCCGAGCGGAGTGCCCCTGGAGT-GA 165
Mycoarachis 116 ---GCCTATGGAAATCCCGAGCAGAGC---------- 139
Graphiola 136 --CGTCTATGCTAATCCCGAGCAGAGTGCGCGAGTAGT-GA 173
C.albicans 154 GTGTATGGGGTTAATCCCGTGGCGAGCCGTCAGGGCGCGAG 194
Lp1 164 ATCTTTCAGGACCCTC-GTAGAGCGCGGTAAGGTGTGGGTG 203
G.graminis 176 CTGGGCGCAGGCCCGT-GTAGAGCGCGGAAAGGTGTGGGCT 215
Beauveria 173 ATCTTCCAGGACGCAT-GTAGAGCGCGGAAAGGTGTGGGTG 212
Pullularia 166 TCCTGGGAACACCCGT-GTAGAGCGCGCCAAGGTGTCGGTG 205
Mycoarachis 140 --CTGTAGAGGCCTGT-GTAGAGCGCGCCAAGGTGTCGGTC 177
Graphiola 174 TACTCGATGCACCTGT-GTAGAGCACGCTAAGGTGTCGGTC 213
C.albicans 195 TTCTGGCAGTGGCCGTCGTAGAGCACGGAAAGGTATGGGCT 235
Lp1 204 GTCTCCTCGGG----------AGACTGCTTAAGGGACGTGC 234
G.graminis 216 CATTTGATGGG----------CTTAAGGGACGTGC 240
Beauveria 213 ACTCTTCTCGGGATGCCTAGAAGGTTGCTTAAGGGACGTGC 253
Pullularia 206 GTGCTCGGGTCATAC-CCGAGACCCACTTAAGGTACGTGC 245
Mycoarachis 178 CGTCCCCG---------CGGGGACGGGCTTAAGGTACGTGC 209
Graphiola 214 GCGCTTGGATGCATAT-CCAAGGCG-GCTTAAGGTACGTGC 252
C.albicans 236 GGCTCTCTG----------AGTCGGCTTAAGGTACGTGC 264
Lp1 235 CAGACC-CACGGGAAACCGTGCCGAATGCG-AAGGACCTGC 273
G.graminis 241 CAAACC-CTCGGGAAACCGAGCCCATGCTCAAGCGCCTGC 280
Beauveria 254 CAGACC-CACGGGAAACCGTGCCGGATGCG-AAGGACCTGC 292
Pullularia 246 CGGCCC-CGATCGAAAGATTGCCTTTAACTAGAGTGCCTAC 285
Mycoarachis 210 CGACCC-CCCGCGAAAGCGGGGCC-TAGGGG---------- 238
Graphiola 253 CGTTCC-CACTTGAAAGAGTGTTTGTAGTCG---------- 282
C.albicans 265 CGTCCCACACGATGAAAAGTGTGCGGTGCAGAATAGTTCCG 305
Lp1 274 --AGTCTGGATCATTCGGAGGGCGATTGAGTGCGGCAGGAA 312
G.graminis 281 --AGCGTAAAGGCATAGGGGGCGTACCGCT--CTCGGTCAA 317
Beauveria 293 --AGTCCAGATCAATCCGCGTGGCTTCGAG-GCCCGGAGAA 330
Pullularia 286 C-ACACGAAGGTTATGGGGGCGTCGA--TTCAAGACGCTGG 323
Mycoarachis 239 ----ATGAGCAGCCGTCGTGCGACGCCCCTGGGGAGCATAC 275
Graphiola 283 ----ACGCAGTCCATCGCTGCGCTAGACTACAAAGAAGTAG 319
C.albicans 306 ACAGAACGAAGCTGCGCCCGAGAAAGCGATTTCTTGGAGCA 346
Lp1 313 ATGCCCGCTACGAGCCTGGTAGAATCACACAGTCAGGGGAT 353
G.graminis 318 ACGAGACG-ACAAGCTTTGGAGGGAAATATCCAAAGCGAGC 357
Beauveria 331 ATGCCCGG-AAGAGCCTGGTACA-CCACATAGATAGGGGAC 369
Pullularia 324 CAGCTCCTAAAAGACATACCTCCCGGCAACCCTGAGGGTCC 364
Mycoarachis 276 GGGCTATGTATAACG---------- 290
Graphiola 320 CCACTTCGGACAAAGAGATTGCACATCTCTTGTAATG---- 356
C.albicans 347 ATG---------- 349
Lp1 354 GAGT-TGTCCTGTGTGGTTGTTCG---------- 377
G.graminis 358 CTGG-TATTCTTCGGAGAATCATTTTTGTTACTAGTCTTCT 397
Beauveria 370 GAGCGTGCGCCGTGCCAGCGGTGGGCGGGCACGCATCATTT 410
Pullularia 365 GTGGAGACAATTGATGGAGGGACACCCAAGTGGTCTCCCNC 405
Mycoarachis - ---------- -
Graphiola - ---------- -
C.albicans - ---------- -

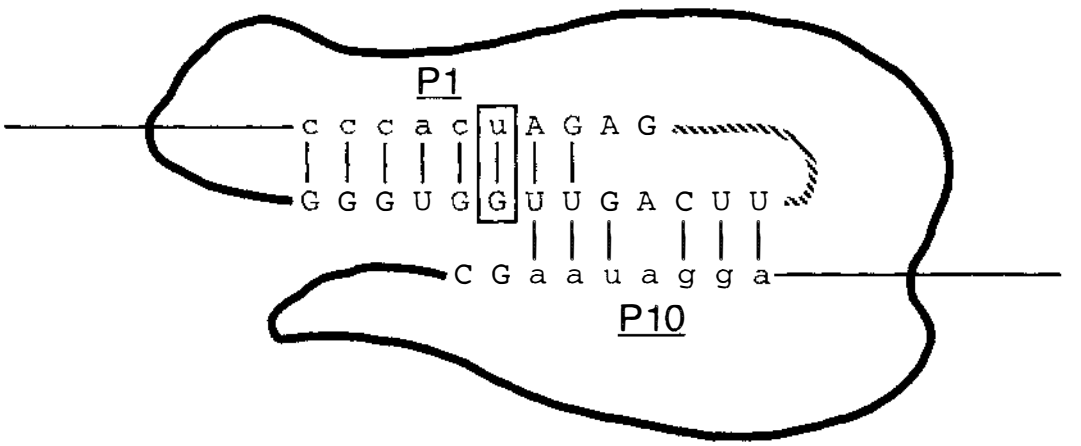



Figure 3.16 Putative internal guide sequence of the Lp1 group I intron

Diagram showing the potential P1 and P10 pairing elements that form the group I intron internal guide sequence which is believed to be important in accurate cleavage and ligation of the exon sequences. The group I intron is shown as a bold line and nucleotides are in upper case. The exons are shown as horizontal thin lines and the nucleotides are in lower case. Pairing is indicated by vertical lines, with G-U pairings allowed. The bold, hatched line indicates that the flanking nucleotides are joined with no nucleotides in between. The box indicates the conserved U-G pairing that occurs at the 5' exon/intron junction. Diagram not drawn to scale, and the bold group I intron line is not indicative of intron structure.

# 3.3 INTERLOCUS HOMOGENISATION OF THE rDNA IN THE HYBRID, Lp1

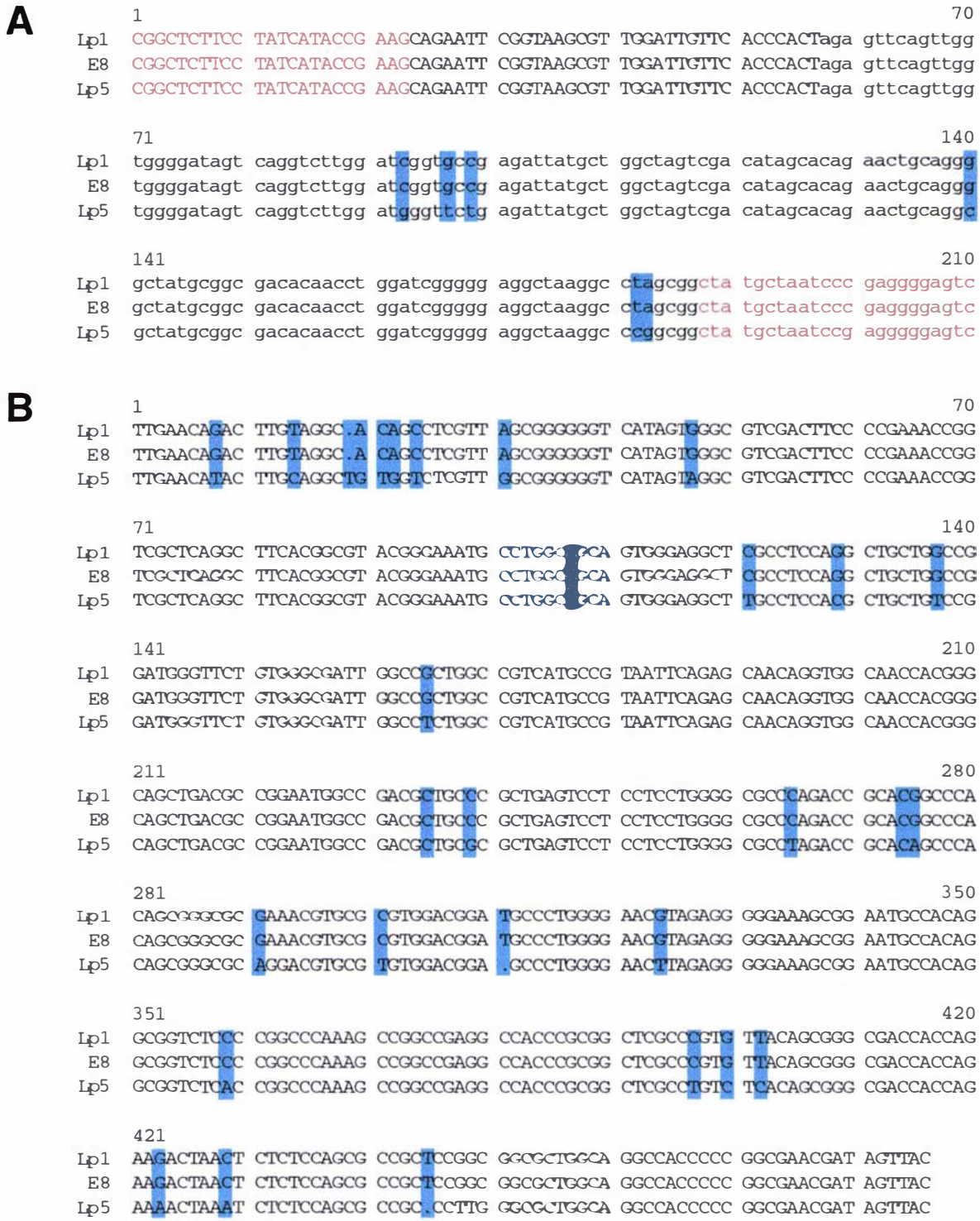
## 3.3.1 Lp1 rDNA IS EXCLUSIVELY DERIVED FROM E8

To determine how the sequence of the rDNA in Lp1 related to its hybrid origin, PCR was performed with genomic DNA from E8 and Lp5 using the primer combinations nts1 - nts2, nts3 - nts4 and nts7 - nts8. The resulting PCR products were sequenced and compared to those from Lp1 (Sections 3.1.6 and 3.2.1). The alignments for the nts1 - nts2 and nts3 - nts4 PCR products are shown in Figure 3.17, and the Genbank Accession numbers are given in the legend. Lp1 and E8 have identical sequences for these PCR products. Comparing the nts1 - nts2 PCR products between Lp1 and Lp5, there were six substitutions over 164 bp (excluding the primer sequences; 96.4% identity). All these mismatches occur in the group I intron. Comparing the nts3 - nts4 PCR products between Lp1 and Lp5, there were 30 substitutions and three indels over 482 bp (93.7% identity).

No product was amplified using the nts7 - nts8 primer combination with Lp5, and the nts7 - nts8 PCR product from E8 is indentical in sequence to that from Lp1. Therefore the alignment is not shown, but the sequence of the PCR products is included in Figure 3.8A. Lack of amplification from Lp5 using the nts7 - nts8 primer combination indicates that the organisation of the IGS in Lp5 differs from that found in Lp1 and E8. Thus the rDNA in Lp1 appears to be exclusively derived from E8, with no evidence of any Lp5 rDNA sequence being found. The Southern blotting data from Figure 3.5 and 3.9 are consistent with this conclusion.

There are two explanations for the presence of rDNA from only one of the progenitors in the hybrid. Either (a) the rDNA array (s) from Lp5 has been lost in the hybrid (Vaughan *et al.*, 1993); or (b) there has been homogenisation between the rDNA arrays from the two progenitors in the hybrid, such that all the Lp5 rDNA sequence has been replaced with E8 rDNA sequence (Hillis *et al.*, 1991; Wendel *et al.*, 1995). This latter process is known as interlocus homogenisation.

Analysis of various genetic loci, isozyme data and microsatellite data show other *Epichloë* endophytes hybrids to be composites of their respective progenitors as well (Christensen *et al.*, 1993; Moon *et al.*, 1999; Tsai *et al.*, 1994). ITS sequences have been determined from these hybrids and in each case only one type of ITS sequence has been found, like Lp1 (Schardl *et al.*, 1991). Therefore presence of only a single rDNA type appears to be a defining feature of the *Epichloë* endophyte hybrids.




## Figure 3.17 Comparison of nts1 - nts2 and nts3 - nts4 PCR products between Lp1, E8 and Lp5

A Alignment of the nts1 - nts2 PCR products from Lp1, E8 and Lp5 genomic DNA. Base pair substitutions and indels are shaded blue. The primers are shown in red text. The group I intron is shown in lower case lettering. The Accession numbers for Lp1, Lp5 and E8 are AF049679, AF049680 and AF049681 respectively.

B Alignment of the nts3 - nts4 PCR products from Lp1, E8 and Lp5 genomic DNA. Base pair substitutions and indels are shaded blue. The primers are excluded from the analysis. The Accession numbers for Lp1, Lp5 and E8 are AF049246, AF049678 and AF049677 respectively.

### 3.3.2 CHROMOSOMAL LOCATION OF THE rDNA

To distinguish between the chromosome loss hypothesis and the interlocus homogenisation hypothesis I looked at the chromosomal location of the rDNA in Lp1 A0 and the two progenitors. The chromosomes were separated using CHEF gel electrophoresis (Figure 3.18A) and range from around 1 Mb to over 5.7 Mb in size as determined by comparison to *S. cerevisiae* and *S. pombe* chromosome size markers (Bio-Rad). The chromosomes from E8 were very poorly resolved, and attempts to obtain acceptable resolution of these chromosomes using a variety of separation conditions were fruitless (results not shown). The gel was Southern blotted and probed with the 5.6 kb *Sal*I coding region probe (Figure 3.18B). A number of hybridising chromosomal bands are present for both Lp1 and Lp5. E8, as expected from the poor resolution of the chromosomes, is not clearly resolved. There are at least five hybridising chromosomal bands in Lp1 and Lp5, and at least two in E8. The signal intensities vary considerably, with chromosomal band I from Lp1 (Figure 3.18B, lane 3) showing a very strong signal. A shorter exposure of this autoradiograph was performed to show the largest hybridising chromosomal bands (band I; Figure 3.18C) which is obscured by band II in Figure 3.18B. The strong signal undoubtedly indicates there are more rDNA units in this chromosomal band, and may result from more than one rDNA-containing chromosome migrating at this position, especially in Lp1 as hybridising bands of the same mobility are seen in both Lp5 and E8.

The high degree of complexity of the rDNA indicated by these results is surprising, as most fungi studied to date have only one or two rDNA loci (e.g. Viaud *et al.*, 1996). In order to rule out hybridisation to non-rDNA sequences producing these results, I repeated these separations, but included intact chromosomal preparations that had been digested with the intron-encoded endonuclease, I-*Ppo*I, from *Physarum polycephalum* (Muscarella and Vogt 1989). This endonuclease has an 15 bp recognition sequence that specifically cleaves a site in the 28S *rrn* gene that is conserved in a wide range of eukaryotes (Muscarella and Vogt 1993). The length of the recognition sequence makes cleavage elsewhere in the genome highly unlikely. The separation of the chromosomal bands and probing of the Southern blot with the 5.6 kb *Sal*I coding region probe are shown in Figure 3.19. In all cases (except one hybridising band in Lp5) digestion with I-*Ppo*I abolishes the high molecular weight hybridising chromosomal bands, and instead there is strong hybridisation to a low molecular weight band, corresponding to a single rDNA unit (Figure 3.19B). In Lp5 one high molecular weight chromosomal band still hybridises after I-*Ppo*I digestion (Figure 3.19B, lane 3). This appears to be an artefact, as in other chromosomal separations of Lp5 with I-*Ppo*I digestion this band does not hybridise (results not shown). Once again the resolution of E8 chromosomal bands is very poor. The length heterogeneity of the IGS in Lp1 can be clearly seen (Figure 3.19B, lane 1).

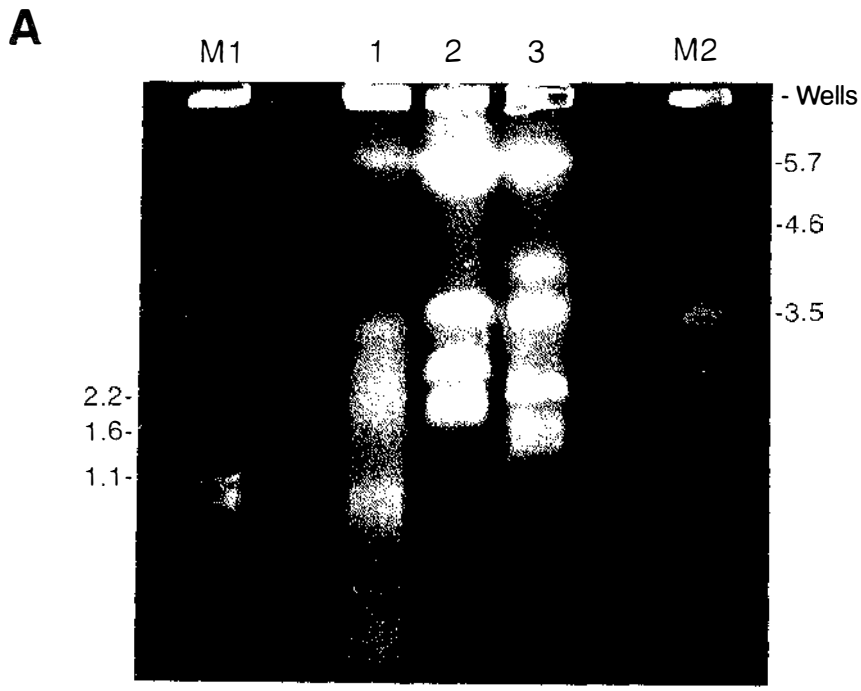
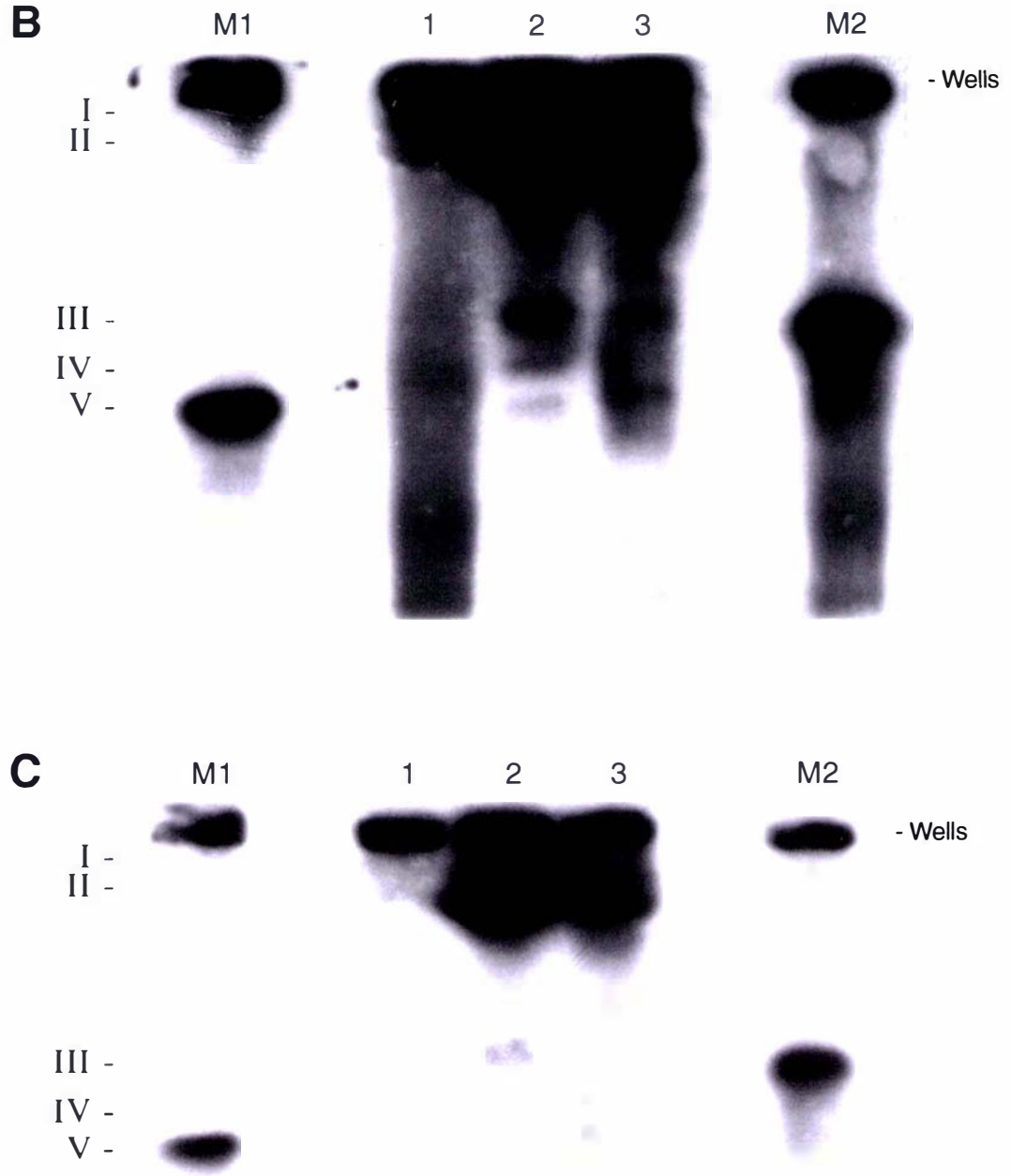



Figure 3.18 The *Epichloë* endophytes contain multiple rDNA clusters

A Ethidium bromide stained 1.0% chromosomal grade agaraose CHEF gel of E8 (lane 1), Lp5 (lane 2) and Lp1A0 (lane 3). See Section 2.19.4 for the CHEF gel electrophoresis conditions. Lane M1 is the *Saccharomyces cerevisiae* chromosome size marker and lane M2 is the *S. pombe* chromosome size marker (both Bio-Rad), and the sizes are indicated in Mb.

B The gel from panel (A) was Southern blotted and probed with the 5.6 kb *Sal*I coding region probe. The chromosomal bands that hybridise in Lp1 are labelled from I to V.

C A shorter exposure of the autoradiograph from panel (B) showing the chromosomal bands I and II in more detail.


B
M1
1
2
3
M2
- Wells
I -
II -
III -
IV -
V -
C
M1
1
2
3
M2
- Wells
I -
II -
III -
IV -
V -

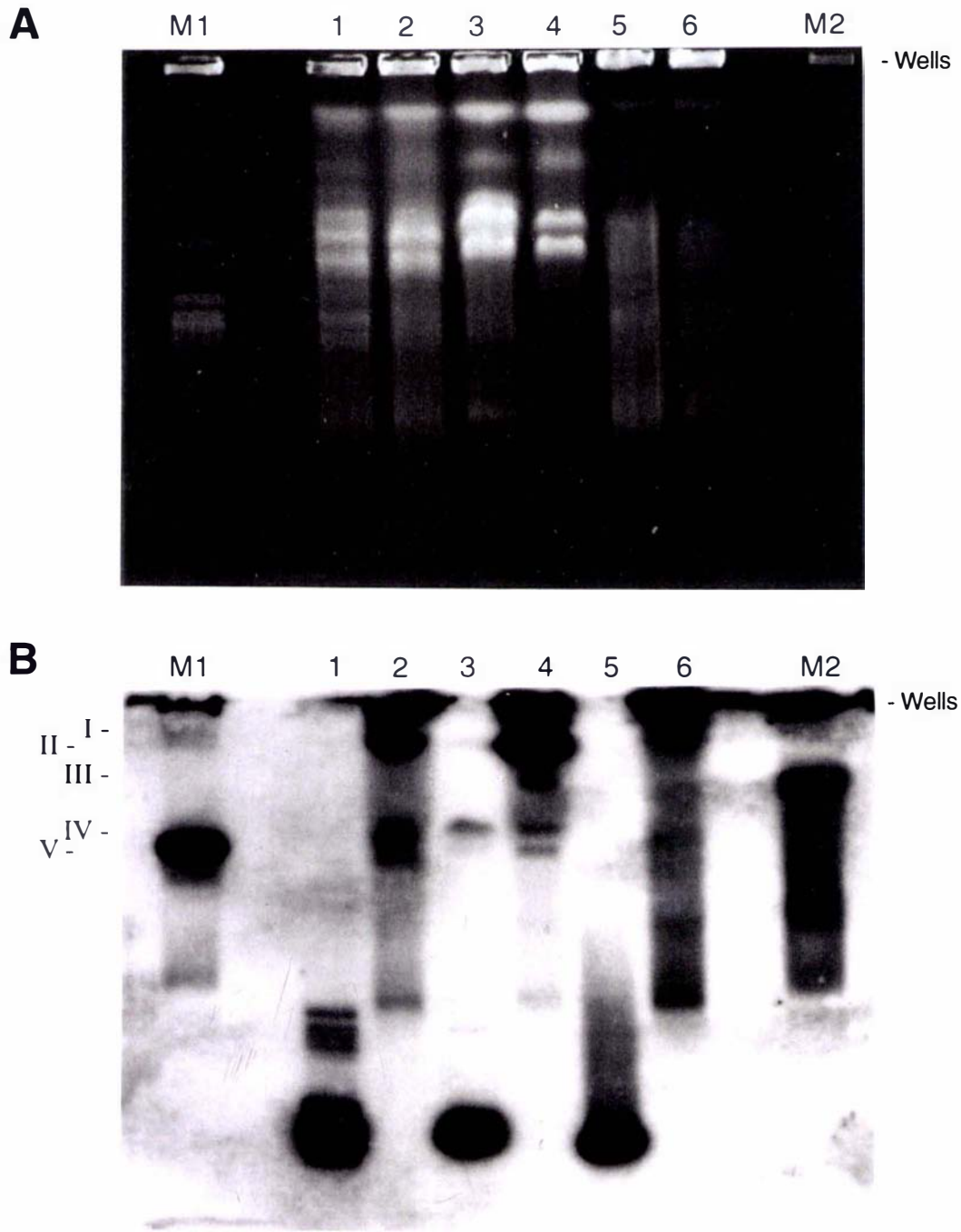



Figure 3.19 Hybridisation of the 5.6 kb *Sal*I coding region probe is specific for the rDNA in CHEF gels

A Ethidium bromide stained 1.0% chromosomal grade agarose CHEF gel of Lp1A0 (lanes 1 and 2), Lp5 (lanes 3 and 4) and E8 (lanes 5 and 6) with the plugs either undigested (lanes 2, 4 and 6) or digested with I-*Ppo*I (lanes 1, 3 and 5). See Section 2.19.4 for the CHEF gel electrophoresis conditions used. Lane M1 is the *S. cerevisiae* chromosome size marker and lane M2 is the *S. pombe* chromosome size marker (both Bio-Rad).

B The gel from panel A was Southern blotted and probed with the nts1 - nts2 PCR product. The chromosomal bands that hybridise in Lp1 are labelled from I to V.

### 3.3.3 SEPARATION OF THE INDIVIDUAL rDNA ARRAYS

The discovery of multiple rDNA loci in these endophytes suggests that several intact rDNA arrays should be able to be resolved on CHEF gels with the appropriate conditions. Therefore intact chromosomal preparations from Lp1A0 were digested with *Bam*HI and *Hin*dIII, two restriction endonucleases that had been previously shown not to cleave within the rDNA (Ganley, 1993 and A.R.D. Ganley, unpublished results). These were separated by CHEF gel electrophoresis and the gel was Southern blotted and hybridised to the 5.6 kb *Sal*I coding region probe. A variety of separation conditions were trialled, and those that gave the best separation of the rDNA arrays are shown in Figure 3.20. There are clearly two high molecular weight rDNA arrays present in the *Bam*H1 digest, and it is possible that the largest band contains more than one rDNA array that migrate together, although no separation conditions were found that could resolve this into more than one band. This large band is at least 1.6 Mb in size, as determined by comparison to the *S. cerevisiae* chromosome standard (Bio-Rad). Also present is a very diffuse smear of hybridisation of lower molecular weight than the well-resolved arrays (Figure 3.20B). These diffuse "bands" are located above the smear of ethidium bromide-stained DNA seen in the *Hin*dIII digest that corresponds to the bulk of the digested DNA but not in the *Bam*HI digest (Figure 3.20A). The reason for the highly diffuse nature of these lower molecular weight bands is not clear, as the *S. cerevisiae* chromosomes are well-resolved in the same region of the gel. The size range covered by this diffuse smear is from less than 200 kb up to 1.1 Mb, as determined by comparison to the *S. cerevisiae* chromosome standard (Bio-Rad). Interestingly, the two high molecular weight rDNA arrays that are resolved in the *Bam*HI digest are not seen at all in the *Hin*dIII digest. This was the case in all the separations that were trialled (results not shown). The reason for this is not clear.

In order to demonstrate that these bands were all actual rDNA arrays, and to extend the analysis to the two progenitors of Lp1, CHEF gel electrophoresis was performed again, this time including E8 and Lp5, and also including intact chromosomal preparations that had been doubly-digested with *Bam*HI and I-*Ppo*I (Figure 3.21A). The gel was Southern blotted and probed with the 5.6 kb *Sal*I coding region probe (Figure 3.21B). As expected, digestion with I-*Ppo*I abolished the hybridising bands seen in the *Bam*HI digest alone, and instead strong hybridisation to a band corresponding to a single rDNA unit was observed. This indicates that both the discrete and diffuse hybridising bands seen in Figure 3.20B are rDNA arrays. The single rDNA unit band can even be seen in the ethidium bromide-stained gel photo (Figure 3.20A). The higher molecular weight discrete hybridising bands have not resolved well for Lp5, so it is not clear whether there are one or two bands. The higher molecular weight discrete hybridising bands from E8 have resolved, and there appear to be two of them. Both Lp5 and E8 also have rDNA arrays that resolve as lower molecular weight diffuse smears (Figure 3.21B).

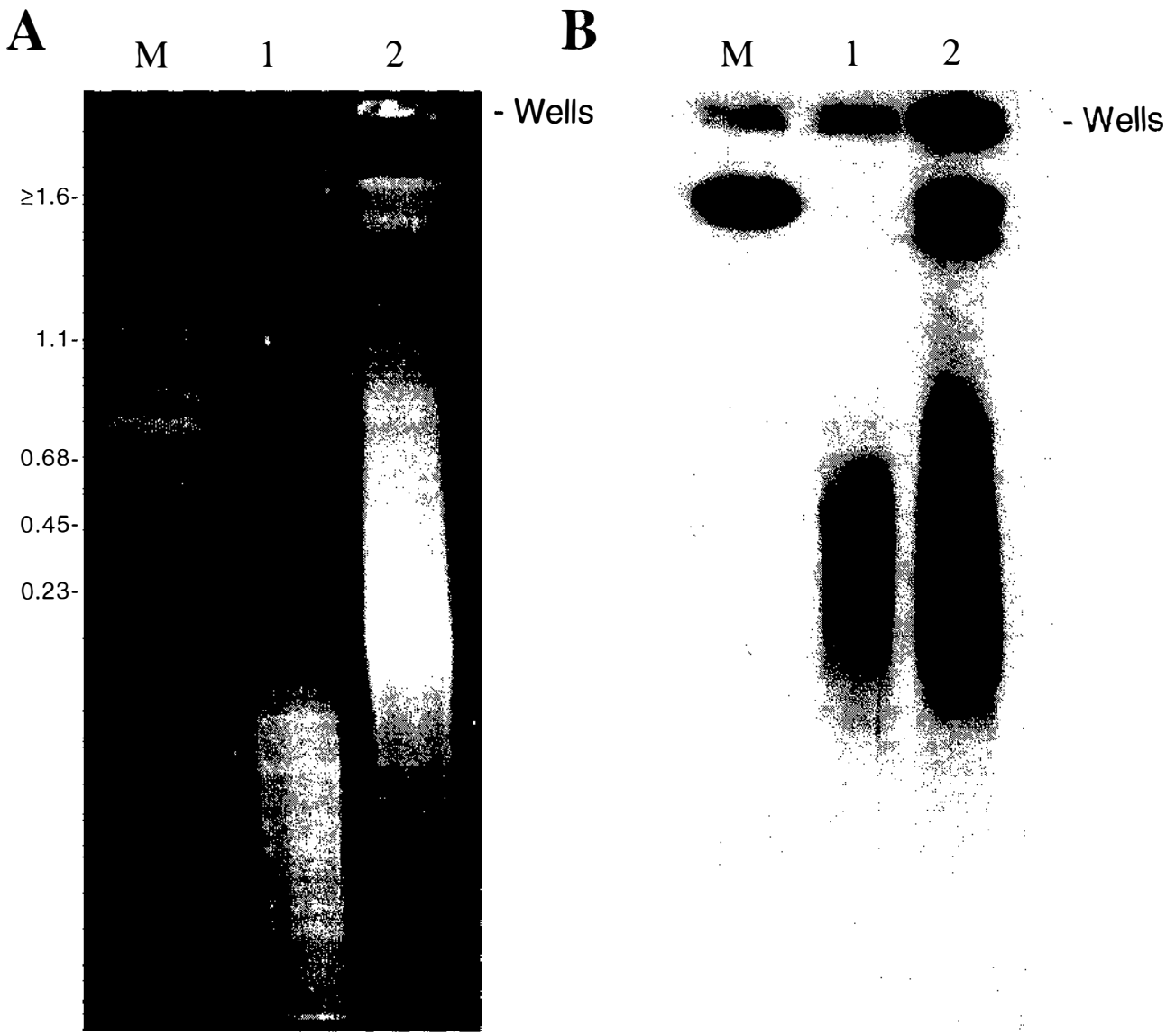



Figure 3.20 CHEF gel separation of intact rDNA clusters by *Hin*dIII and *Bam*HI digestion

A Ethidium bromide stained 1.0% chromosomal grade agarose CHEF gel of Lp1A0 digested with *Hin*dIII (lane 1) and *Bam*HI (lane 2). See Section 2.19.4 for CHEF gel electrophoresis conditions. Lane M is the *S. cerevisiae* chromosome size marker (Bio-Rad). Sizes are indicated in Mb.

B The CHEF gel from panel A was Southern blotted and probed with the 5.6 kb *Sal*I coding region probe.

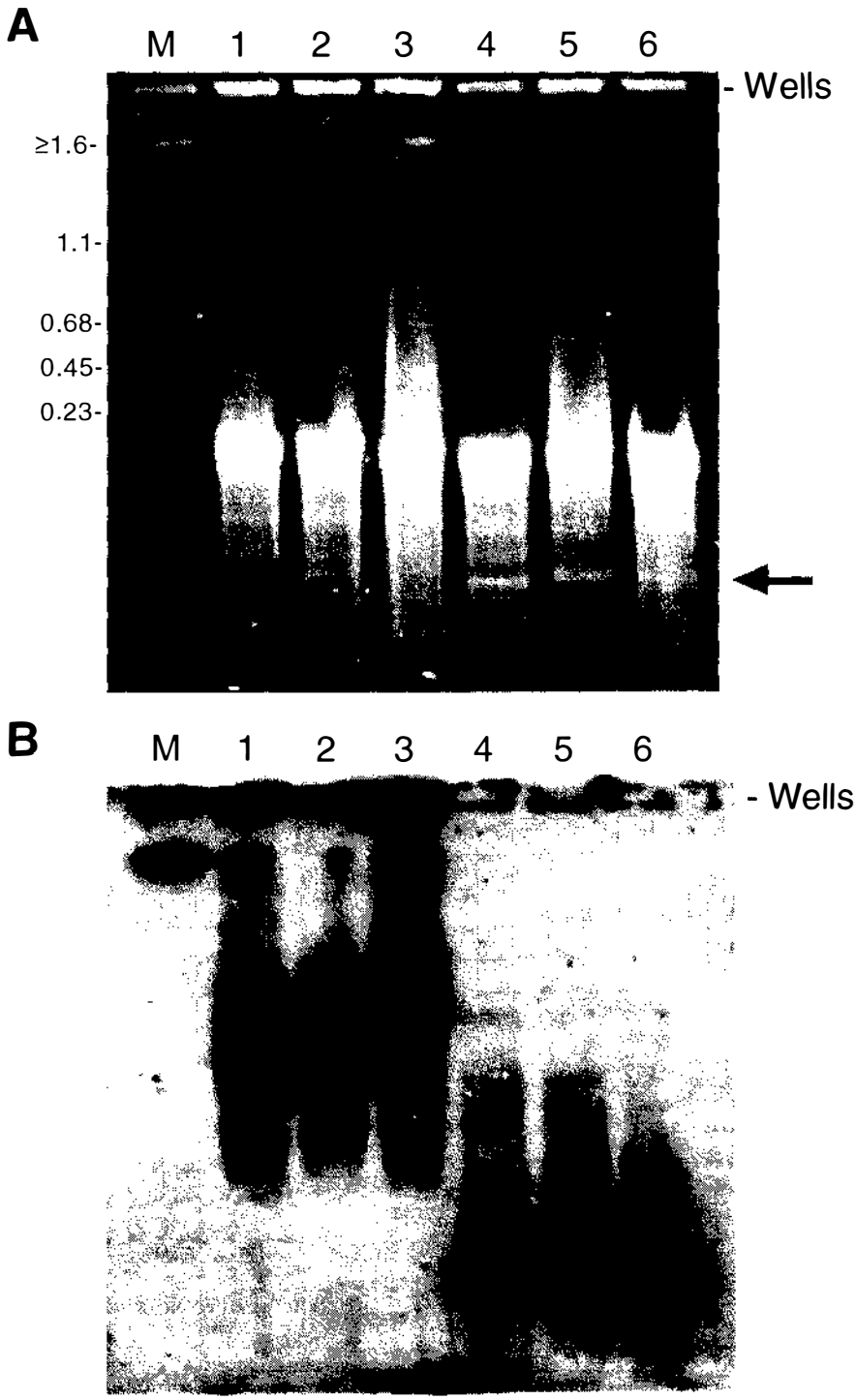



Figure 3.21 CHEF gel separation of rDNA clusters in Lp1, Lp5 and E8

A Ethidium bromide stained 1.0% chromosomal grade agarose CHEF gel of E8 (lanes 1 and 4), Lp5 (lanes 2 and 5) and Lp1 (lanes 3 and 6) digested with *Bam*HI (lanes 1 to 3) and *Bam*HI/I-*Ppo*I (lanes 4 to 6). See Section 2.19.4 for CHEF gel electrophoresis conditions. The prominent band the size of a single rDNA unit is indicated by an arrow. Lane M is the *S. cerevisiae* chromosome size marker (Bio-Rad). Sizes are indicated in Mb.

B The gel from panel A was Southern blotted and probed with the 5.6 kb *Sal*I coding region probe.

### 3.3.4 STRATEGY FOR CLONING rDNA-DNA JUNCTION FRAGMENTS

The results from the previous section were not able to conclusively demonstrate whether the presence of only one type of rDNA sequence in Lp1 was the result of interlocus homogenisation or rDNA array loss. In particular, the inability to adequately resolve the E8 chromosomes prevented a comparison of the rDNA organisation between the hybrid and the two progenitors. The very complex organisation of the rDNA in Lp1 and Lp5 along with the similarities in the chromosomal bands that hybridise to the rDNA in Lp1 and Lp5 suggest that Lp1 does still contain some of the rDNA loci that were derived from Lp5. To try and resolve this further, and additionally to gain some information about the mechanism (s) of homogenisation, I undertook to clone the junction sequences between the rDNA and the rest of the chromosome. Sequence information from the DNA flanking the rDNA has the potential to shed light on both of these problems.

The repetitive nature of the rDNA makes cloning of junction fragments problematic. Assuming 100 rDNA units per array, there would be at least a 50-fold excess of rDNA-only fragments over rDNA junction fragments in a digest. To obviate screening large numbers of positive clones from a library, I decided to use a PCR-based cloning method. To reduce problems caused by a large excess of non-junction sequences, I used a biotin-streptavidin enrichment procedure (making use of magnetic bead technology) to enrich for junction sequences. The methodology for the cloning of rDNA junction sequences is set out in Figure 3.22, as well as in Section 2.20. Briefly, genomic DNA was digested with a restriction endonuclease that does not cleave the rDNA (*Bam*HI and *Hin*dIII were used) and the high molecular weight DNA containing the rDNA arrays was gel extracted from a Seaplaque agarose gel. A composite "linker" was designed. This is composed of two regions. The first is a priming site for the TAG primer that was used in the multi-variant repeat PCR in Section 3.1.8. The second is a multi-cloning site that includes restriction sites for *Hin*dIII and *Bam*HI. This linker is also tagged with a 5' biotin label (at the TAG priming site end of the linker). The linker was digested with the same restriction endonuclease as the genomic DNA, and the genomic DNA and the linker were ligated together. This ligation mixture was digested with I-*Ppo*I to cleave the rDNA units and mixed with the streptavidin-labelled Dynabeads, which were separated with a magnetic separator. Theoretically no rDNA fragments except those including the junctions will have a linker ligated. Therefore only junction fragments, along with non-rDNA fragments that have also ligated to the linker, will have a biotin label and will remain after the streptavidin-labelled Dynabead magentic separation. From here PCR was performed using the Dynabeads as the template with a primer to some part of the rDNA and the TAG primer. Nested PCR was then performed on this original PCR reaction using an rDNA primer adjacent to the first rDNA primer, and the TAG primer again. Amplified bands were cloned and analysed.

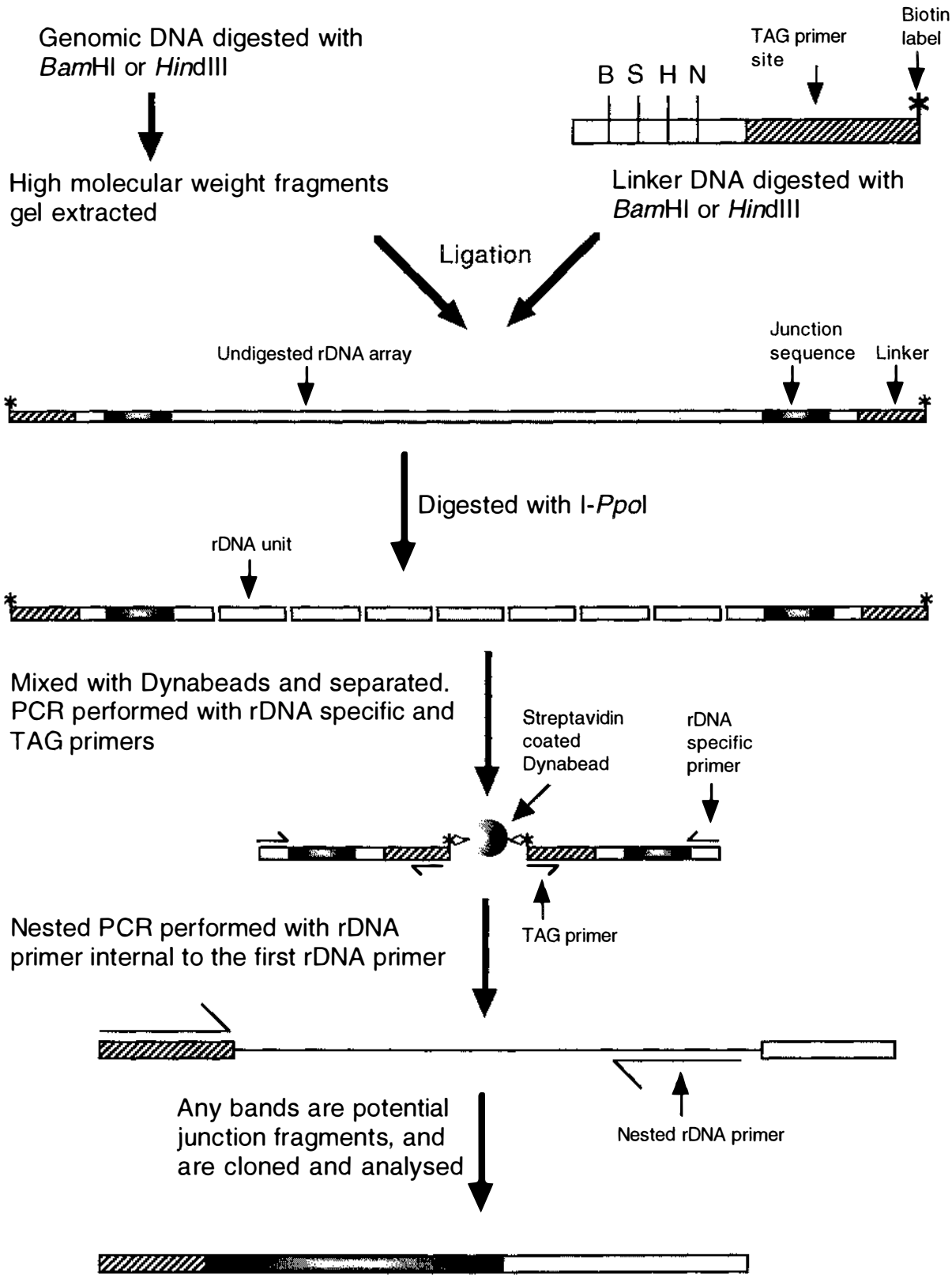



Figure 3.22 PCR method for cloning rDNA junction fragments

See text for details. Ribosomal sequence is stippled, the TAG primer sequence is hatched, the linker polycloning site is boxed in white and rDNA junction sequence is gradient-shaded. B, S, H and N indicate *Bam*HI, *Sma*I, *Hin*dIII and *Not*I sites respectively in the linker. Diagram not drawn to scale.

### 3.3.5 RESULTS OF THE rDNA JUNCTION CLONING

Initially this system was trialled using pUC118 digested with *Bam*HI or *Hin*dIII as a control, instead of genomic DNA. PCR was carried out using the M13-forward or M13-reverse primers, and the TAG primer. Bands of the expected size were present for both the M13-forward and M13-reverse primers with the *Hin*dIII digest (Figure 3.23A). Bands of the expected size were also present for both primers with the *Bam*HI digest, but for both primers an additional band, about 20 bp larger than the expected product, was also seen. This result is reproducible for the *Bam*HI digest (results not shown), but has never been seen with the *Hin*dIII digest. It is not known where this additional band comes from. However these results indicate that the system does work.

I then carried out this junction cloning method using Lp5 and E8 genomic DNA, doing parallel reactions with *Bam*HI and *Hin*dIII. I decided to clone the junctions from the progenitors of Lp1, as once junction fragments were obtained the situation in the hybrid could easily be ascertained through the use of primers designed to the junction sequences. An example of typical results obtained through the course of the junction cloning are shown for *Hin*dIII in Figure 3.23. The first round of PCR amplification was carried out after digestion, ligation and Dynabead enrichment using primers nts1 and nts3 to clone junctions out one side of the rDNA array, and primers nts2 and nts4 to clone junctions on the other side of the rDNA array (Figure 3.8C). These amplifications routinely produced a smear of amplification products - discrete bands were sometimes observed, but never without an accompanying smear (Figure 3.23B), and in conditions where the smear was not observed, no amplification was evident. Nested PCR amplification was performed using the nested primers nts11, nts12, nts13 and nts14 (refer to Figure 3.8A) on the first-round PCR amplifications with nts1, nts2, nts3 and nts4 respectively. Discrete products were often observed from this nested PCR (e.g. Figure 3.23C, lane 1), although once again with an accompanying smear, and they did not correspond to the discrete products occasionally seen in the first-round PCR amplification. Stabbed band PCR was used to amplify individual bands (as often several bands per reaction were observed), and when a single band was produced (Figure 3.23D) these were cloned into a pGEM-T vector. Several of the discrete bands observed were present regardless of the nts primer used, and therefore these were not cloned as they are presumed to result from non-specific annealing of the TAG primer. For the *Bam*HI digest, I could only get stable stabbed band PCR reamplification for 2.0 kb and 2.5 kb products from Lp5 with nts4/nts14 primer set. For some unknown reason I was not able to clone these bands into the pGEM-T vector (results not shown). For the *Hin*dIII digest I could stably reamplify 2.2 kb, 2.7 kb and 5.5 kb bands from E8 and 3.2 kb and 4.4 kb bands from Lp5, all with the nts1/nts11 primer set (Figure 3.23D). I cloned all of these except the 5.5 kb E8 band (for which I did not get any transformants) into the pGEM-T vector. I was unable to stably reamplify any bands with the nts2/nts12, nts3/nts13 or nts4/nts14 primer sets using stabbed band PCR reamplification from the *Hin*dIII digest.

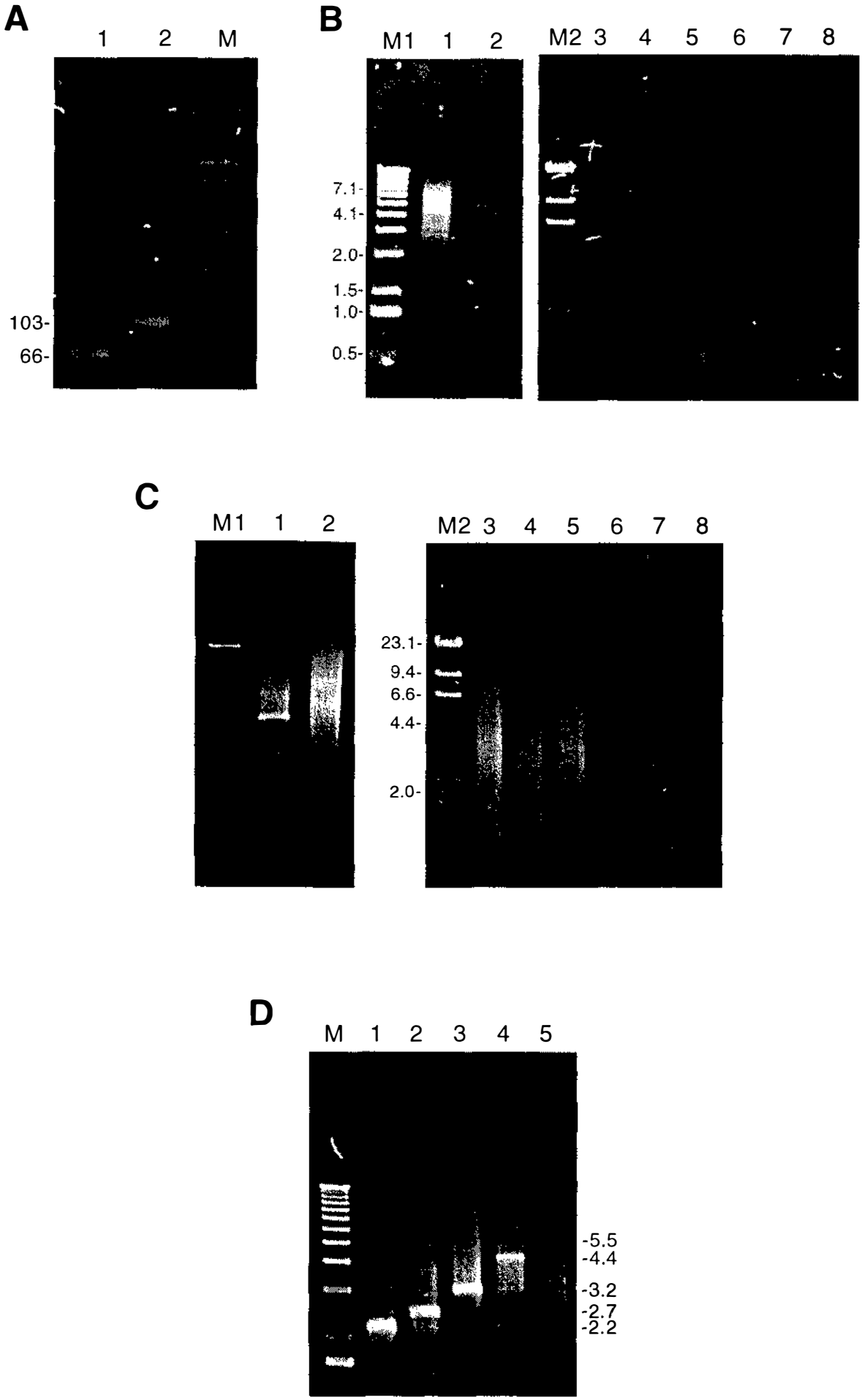
## Figure 3.23 rDNA junction fragment PCR cloning using *Hin*dIII

A Ethidium bromide stained 3% NuSieve agarose gel showing the positive control PCR products of the linker DNA ligated to pUC118. The fragments were amplified using the TAG primer and the M13 forward primer (66 bp; lane 1) and M13 reverse primer (103 bp; lane 2) from Dynabead-enriched template DNA, and are indicated. Lane M is the 100 bp size marker (Gibco BRL).

B Ethidium bromide stained 0.7% agarose gels showing the first round of PCR amplification. PCR was performed with the TAG primer and nts1 (lanes 1 and 2), nts2 (lanes 3 and 4), nts3 (lanes 5 and 6) and nts4 (lanes 7 and 8) with *Hin*dIII-digested, Dynabead-enriched, E8 genomic DNA (lanes 1, 3, 5 and 7) and Lp5 genomic DNA (lanes 2, 4, 6 and 8). Lane M1 is the 1 kb ladder size marker (Gibco BRL) and the sizes are shown in kb; lane M2 is the λ *Hin*dIII size marker.

C Ethidium bromide stained 0.7% agarose gels showing the nested PCR amplification. Lanes are as in panel (B) except the nested primers nts11, nts12, nts13 and nts14 were used in place of nts1, nts2, nts3 and nts4 respectively, and the PCR templates were the respective 500-fold dilutions of the reactions from panel (B). The sizes of the λ *Hin*dIII ladder markers are shown in kb.

D Ethidium bromide stained 0.7% agarose gel showing PCR amplified potential junction fragments. All fragments were amplified from the TAG - nts11 nested PCR product by stabbed band PCR reamplification. Lanes 1, 2 and 5 are from E8 genomic DNA and lanes 3 and 4 are from Lp5 genomic DNA. Sizes of the amplified bands are indicated in kb. Lane M is the 1 kb ladder size marker.


A
1
2
M
103-
66-
B
M1
1
2
M2
3
4
5
6
7
8
7.1-
4.1-
2.0-
1.5-
1.0-
0.5-
C
M1
1
2
M2
3
4
5
6
7
8
23.1-
9.4-
6.6-
4.4-
2.0-
D
M
1
2
3
4
5
-5.5
-4.4
-3.2
-2.7
-2.2

The cloned PCR products could all be tested directly for junction sequences by PCR, as primer nts2 is located close to the nested primer, nts11, and in the opposite direction (Figure 3.8A). PCR amplification using nts2 should give a PCR product if the clone is a true junction fragment, unless the junction occurs in the ~ 100 bp between the two priming sites. In all cases no product could be amplified using nts2 and the M13-forward and M13-reverse primers, suggesting these clones were not true junction fragments (results not shown).

Despite these cloned bands appearing not to be true junction fragments, I decided to sequence one of them. The 3.2 kb fragment from *Hin*dIII-digested Lp5 genomic DNA amplified with the nts1/nts11 nested primer set was chosen, and was sequenced in from both ends to determine its origin. Surprisingly, the sequence from the nested primer, nts11, was very similar to the Lp1 IGS sequence from the same region, as expected for a junction clone. The sequence in from the TAG primer showed no similarity to the Lp1 IGS sequence, nor to any sequence from the Genbank database. The sequence in from the nts11 and TAG primers is shown in Appendix 2. The 3.2 kb nts1/nts11 PCR product appears to be a true junction fragment, although because of the divergence of the Lp5 IGS from that of Lp1 (Section 3.3.1), it cannot be ruled out that this is actually Lp5 IGS sequence. Unfortunately time constraints, primarily due to significant difficulties in initially obtaining amplification products from the Dynabeads, meant I was unable to characterise this potential junction fragment further. It is not known why the PCR amplification using nts2 and the M13 primers failed. On the basis of this result the other bands that were cloned should be sequenced, as the PCR does not appear to be a reliable test and they may well be junction fragments too.

## 3.4 CHARACTERISATION OF THE 5S RIBOSOMAL RNA GENES

Wendel *et al.* (1995) showed that interlocus homogenisation had occurred in the rDNA of a cotton hybrid, but Cronn *et al.* (1996) found that the 5S rRNA gene arrays had been maintained as their original progenitor sequences. Other studies looking at the 5S rRNA genes from hybrids indicated that the 5S rRNA arrays from the progenitors remain stable in the hybrid (Dvorák, 1989; Kellogg and Appels, 1995; Reddy and Appels, 1989; Sastri *et al.*, 1992). This led me to hypothesise that interlocus homogenisation is much more frequent between rDNA arrays than between 5S rRNA arrays, and in hybrids the 5S rRNA arrays should maintain evidence of a hybrid origin. To test this hypothesis I decided to clone the 5S rRNA genes from Lp1, Lp5 and E8 to see if Lp1 maintained both Lp5 and E8 5S rRNA arrays in its genome.

### 3.4.1 CLONING THE 5S rRNA GENES

PCR primers were designed to amplify 5S rRNA genes from a tandem array as set out in Figure 3.24A. The primers were designed from an alignment of 5S rRNA sequences from pyrenomycetous fungi found in the Genbank database. Primer 5S1 was designed (5' to 3') to

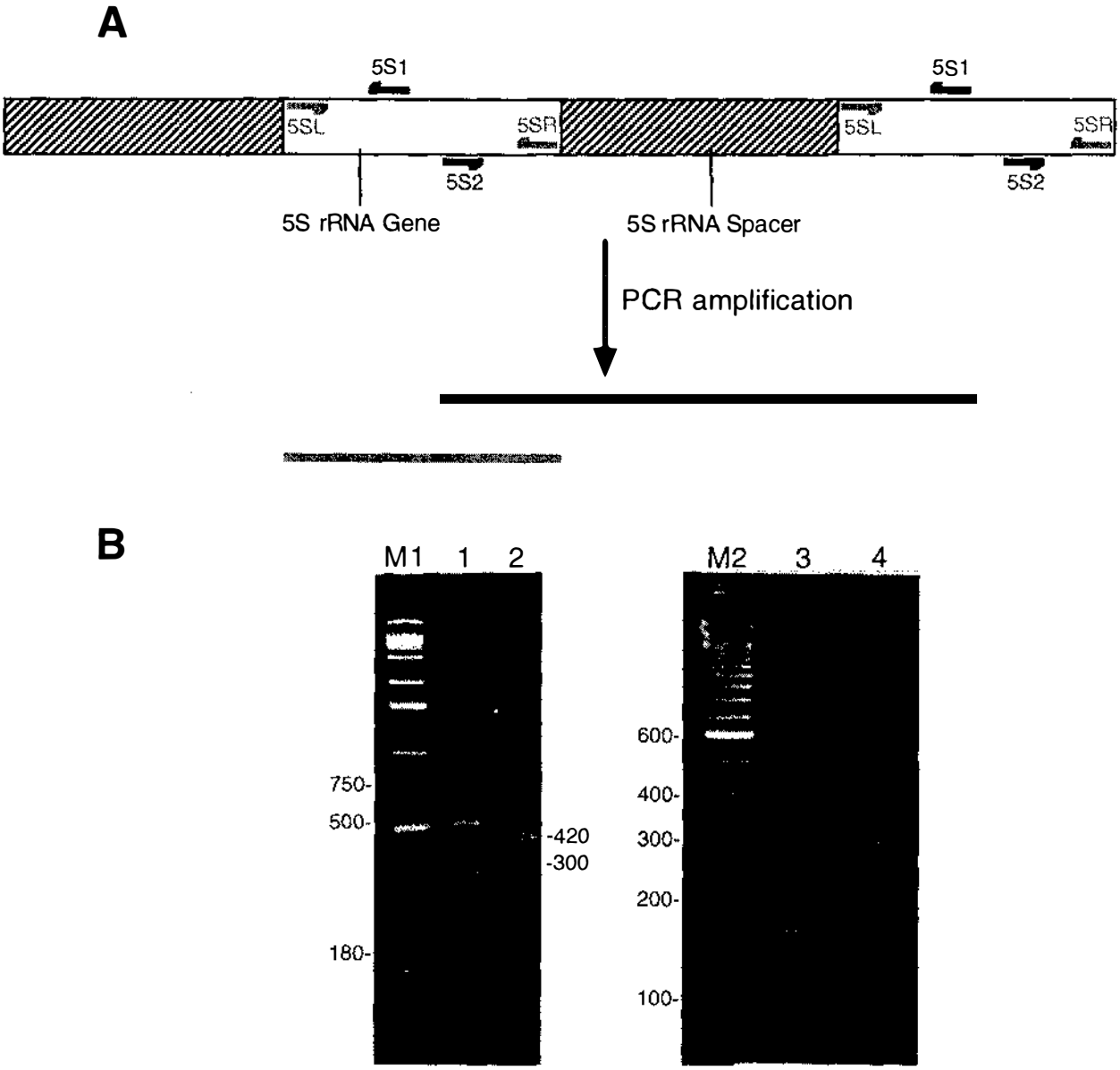



Figure 3.24 Initial PCR cloning of the 5S rRNA genes

A Stylised diagram outlining the strategy for cloning the 5S rRNA genes. Primers 5S1 and 5S2 (in black) were designed assuming a tandem arrangement of the 5S rRNA genes. PCR amplification would give a PCR product containing the ends of the 5S rRNA gene and a 5S rRNA spacer (black, heavy line). Primers 5SL and 5SR (shaded) were designed for internal 5S rRNA gene PCR amplification, and their predicted product (shaded, heavy line) is also shown. See Table 2.2 for primer sequences.

B Results of PCR amplification using primers 5S1 and 5S2 with Lp5 (lane 1), E8 (lane 2), Lp1O (lane 3) and Lp1D0 (lane 4) genomic DNA. The sizes of the bands cloned and sequenced are indicated in bp and lane M1 is the 1 kb ladder size marker (Gibco BRL). Lane M2 is the 100 bp ladder size marker (Gibco BRL), and sizes are indicated in bp. See Table 2.3 for the PCR thermocycling conditions used.

nucleotides 51 to 35 (with the 5' end of the *Acremonium persicinum* 5S rRNA gene, Accession number X00864, numbered as 1), and 5S2 was designed to nucleotides 76 to 92. PCR amplification was performed on Lp1, Lp5 and E8 genomic DNA using the conditions set out in Table 2.3, although initially higher annealing temperatures were used. No products were obtained at these higher temperatures, and therefore the annealing temperature had to be dropped to 47°C to obtain any products. At this annealing temperature a number of weak products were obtained (Figure 3.24B).

I cloned and sequenced two of the products from the E8 5S1 - 5S2 amplification (~ 420 bp and ~ 300 bp), and three from the Lp5 5S1 - 5S2 amplification (~ 500 bp, ~ 220 bp, and ~ 180 bp). The Lp5 180 bp PCR product was the only one of these PCR products to show any homology to 5S rRNA genes outside of the primer sequence. However the only homology was to the 3' end of a 5S rRNA gene - there was no homology to the 5' end of the next gene at the other end of the clone as expected. I probed a Southern blot of a *Sal*I genomic digest of Lp1 A0, Lp5 and E8 DNA with this PCR product (Figure 3.25). The hybridisation signal was low, but there are a number of hybridising bands - at least 13 in E8, the lane with the strongest signal (Figure 3.25, lane 3). This hybridisation pattern, the difficulty in amplifing 5S rRNA genes and the lack of evidence for tandem organisation from the one positive PCR clone obtained all suggest that the 5S rRNA genes are not organised in a tandem array as they are in most organisms (Hadjiolov, 1984), but instead are dispersed. The origin of the sequence from the other PCR products is not clear, as they do not contain any 5S rRNA sequence, nor do they show homology to anything else in the Genbank database.

### 3.4.2 5S rRNA GENES IN THE *EPICHLOË* ENDOPHYTES ARE DISPERSED

To directly demonstrate that the 5S rRNA genes in these endophytes are dispersed rather than organised in a tandem array, I designed primers to each end of the 5S rRNA gene (primers 5SR and 5SL; Figure 3.24A), using the alignment of 5S rRNA sequences from pyrenomycetous fungi again, as well as sequence from the 180 bp Lp5 5S1 - 5S2 PCR product. The lack of nucleotide conservation at the ends of the 5S rRNA gene made primer design difficult, and necessitated including some degenerate positions in primer 5SL (Table 2.2).

PCR performed with these primers resulted in amplification products of the correct size (117 bp), although amplification for E8 was not as strong as for Lp1 and Lp5 for unknown reasons (Figure 3.26A; see below). These PCR products were directly sequenced, and this confirmed that they are 5S rRNA gene sequences. A sequence alignment of the sequences from Lp1, Lp5 and E8 is shown in Figure 3.26B. The primer sites, which account for 34 of the 117 bp, are excluded from this alignment as they will be identical regardless of the actual genomic sequence. 5S rRNA gene sequences from *Neurospora crassa*, *Acanthamoeba castellani*, *Xenopus laevis* and *Rattus norvegicus* are included in the alignment for comparison. As can be seen, the sequences for the three *Epichloë* endophytes are virtually identical. There are three positions of ambiguity in the

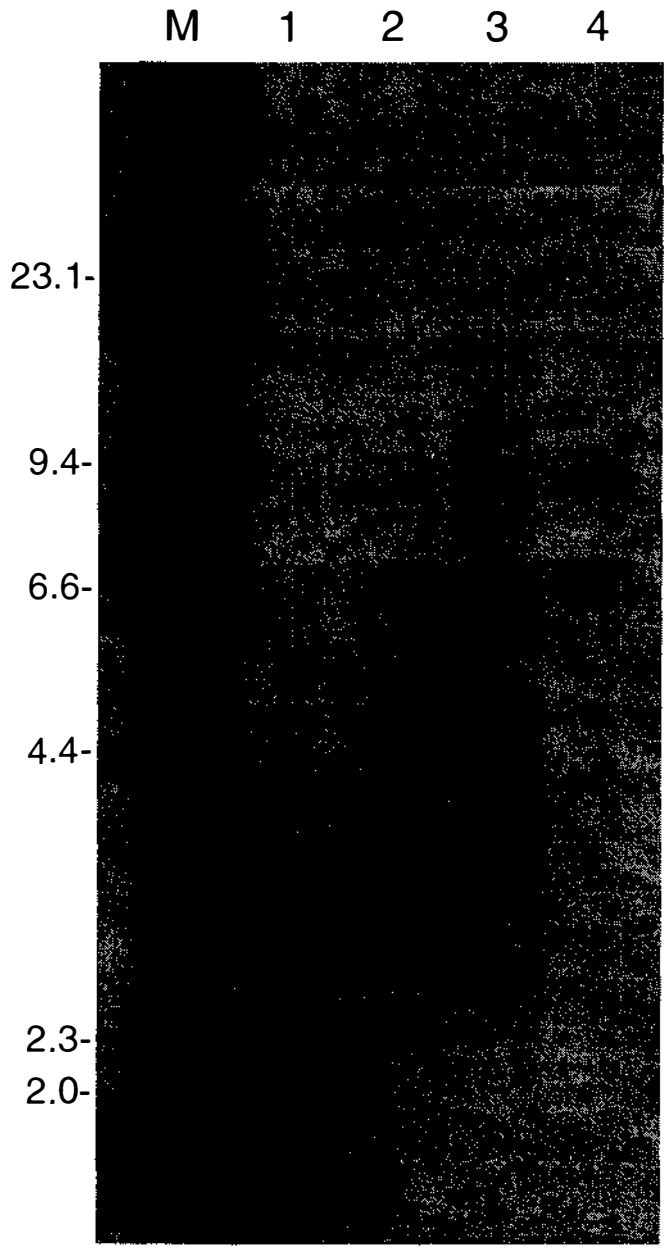



Figure 3.25 Hybridisation of the Lp5 5S1 - 5S2 PCR product to Lp1 and progenitors

1.2 µg of genomic DNA from Lp1O (lane 1), Lp5 (lane 2), E8 (lane 3) and Lp1 A0 (lane 4) was digested with *Sal*I and fractionated on a 0.7% agarose gel. The resulting Southern blot was probed with the 180 bp 5S1 - 5S2 PCR product from Lp5. Lane M is the λ *Hin*dIII size markers, and sizes are indicated in kb.

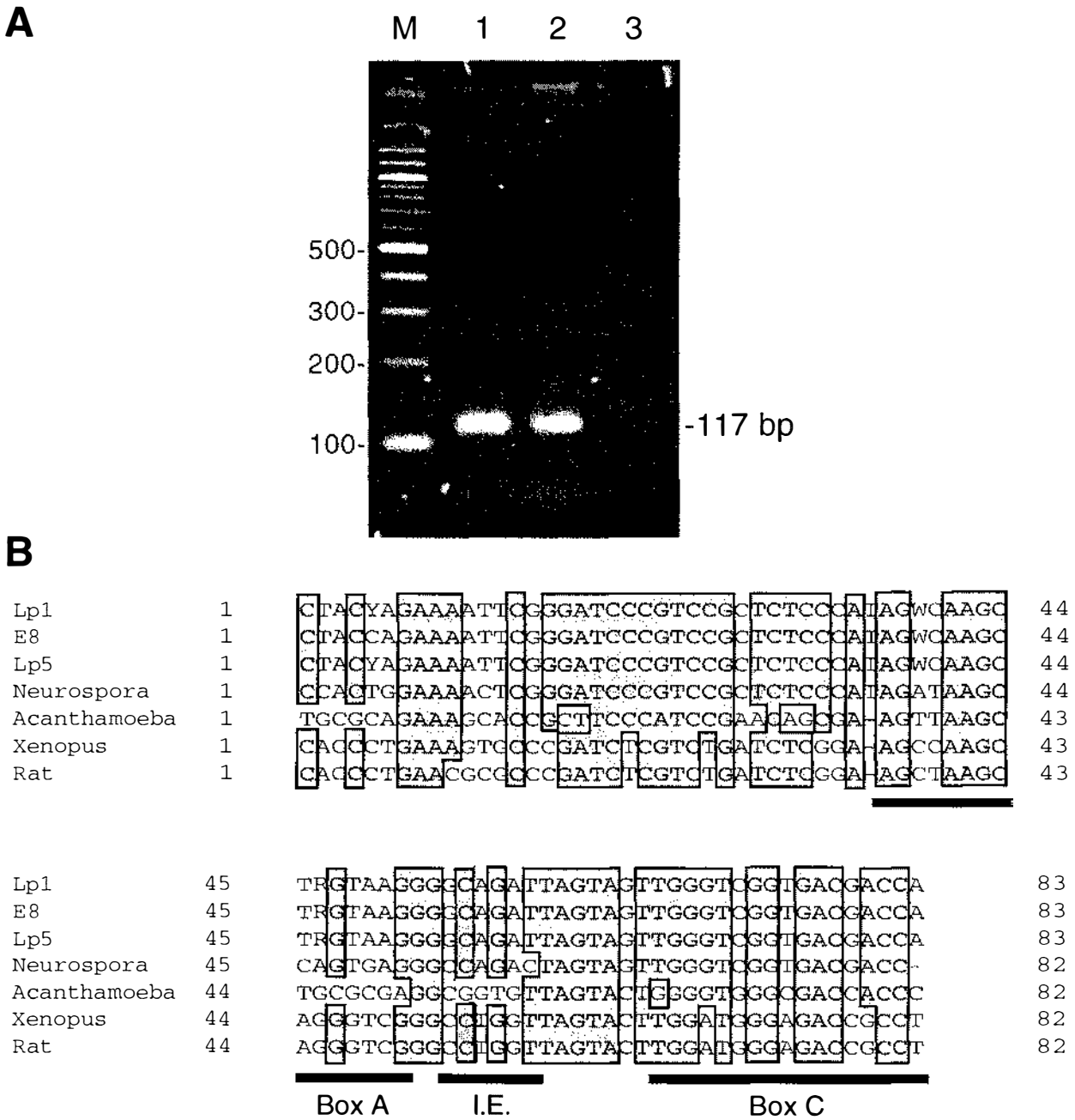



Figure 3.26 Amplification and characterisation of *Epichloë* endophyte 5S rRNA genes

A Ethidium bromide stained 3% NuSieve agarose gel showing the PCR amplification of the 5S rRNA genes from Lp1A0 (lane 1), Lp5 (lane 2) and E8 (lane 3) genomic DNA using primers 5SL and 5SR. The 117 bp product is indicated. Lane M is the 100 bp ladder size marker (New England Biolabs) and sizes are indicated in bp.

B Alignment of the amplified portion of the 5S rRNA genes from Lp1A0, Lp5 and E8. Also shown aligned are the equivalent portion of the 5S rRNA genes from *Neurospora crassa*, *Acanthamoeba castellani*, *Xenopus laevis*, and *Rattus norvegicus* (Accession numbers K02952, M28984, J01899 and K01594 respectively). Box A, the intermediate element (I.E.) and box C are indicated with heavy underlines. See text for details. Nucleotides conserved in 5 of the 7 genes are boxed in grey. The ambiguity codons are: Y = C or T; W = A or T; and R = A or G.

sequences. Two of these, at positions 39 and 46, are shared between all three endophytes, and the other is shared by Lp1 and Lp5 at position 5. As these are sequenced directly from PCR products they are effectively a consensus sequence of all the 5S rRNA genes that amplify from each genome. These ambiguous positions are likely to represent polymorphisms present between the different copies of the 5S rRNA genes. The poor amplification of E8 with 5SR and 5SL may be the result of base-pair mismatches at the priming site, however the three endophytes have nearly identical sequences for the amplified portion of the 5S rRNA gene. The positions of the A and C boxes and the intermediate element are indicated in Figure 3.26B (Pieler *et al.*, 1987).

To demonstrate the dispersed nature of the 5S rRNA genes in these *Epichloë* endophytes, genomic DNA from Lp1, Lp5, E8 and also *Epichloë festucae* isolate Frc7 was digested with *Bam*HI, *Bgl*II, *Eco*RI and *Sma*I, separated on an agarose gel and the Southern blot probed with the 5SR - 5SL PCR product from Lp1 (Figure 3.27). *E. festucae* was included, as *N. lolii* LpTG-1 (which includes Lp5) is believed to be derived from it (Schardl *et al.*, 1994), and there has been some debate as to whether *N. lolii* LpTG-1 might be an ancient hybrid (M. Christensen, personal communication). The results clearly demonstrate that the 5S rRNA genes in all these endophytes are dispersed rather than organised in a tandem cluster. The hybridisation strength of the bands is fairly uniform, with many bands displaying a similar hybridisation intensity. This indicates that many 5S rRNA genes are present as a single copy per locus (or at least a standard number of copies per locus). Furthermore Lp1 contains bands from both Lp5 and E8, suggesting additivity in the 5S rRNA gene profile in the hybrid. A few of the bands in Lp1 are novel, and several from both Lp5 and E8 are not present in Lp1 demonstrating that some changes have occurred upon or subsequent to the hybridisation. The banding patterns of Lp5 and *E. festucae* are very similar, although again there are some changes. The similarity in the 5S rRNA gene profiles of *E. festucae* and *N. lolii* LpTG-1 argues against *N. lolii* LpTG-1 being a hybrid.

### 3.4.3 THE rDNA AND 5S rRNA GENES ARE LOCATED ON THE SAME CHROMOSOMAL BANDS

I reprobed the CHEF Southern blot shown in Figure 3.18B with the 5SR - 5SL PCR product from Lp1 to determine the dispersion of the 5S rRNA genes in the genomes of these *Epichloë* endophytes (Figure 3.28A). The Southern blot probed with the 5.6 kb *Sal*I coding region probe from Figure 3.18B is reproduced here for comparison (as Figure 3.28B). Surprisingly, exactly the same chromosomal bands hybridise with the rDNA and 5S rRNA probes. The hybridisation signals for the 5S rRNA probe are much more uniform than those with the rDNA probe, indicating that approximately equal numbers of 5S rRNA genes are present on each chromosomal band. The *S. cerevisiae* and *S. pombe* 5S rRNA genes fail to hybridise, presumably because the 5S rRNA genes are too dissimilar to cross-hybridise at the stringency used. Hybridising chromosomal band V in Lp1 (Figure 3.28A, lane 3) is only very weakly hybridising, so it is not clear whether this chromosomal band really does contain any 5S rRNA genes. However the same band is also very weakly hybridising with the rDNA probe in Lp1 (Figure 3.28B).

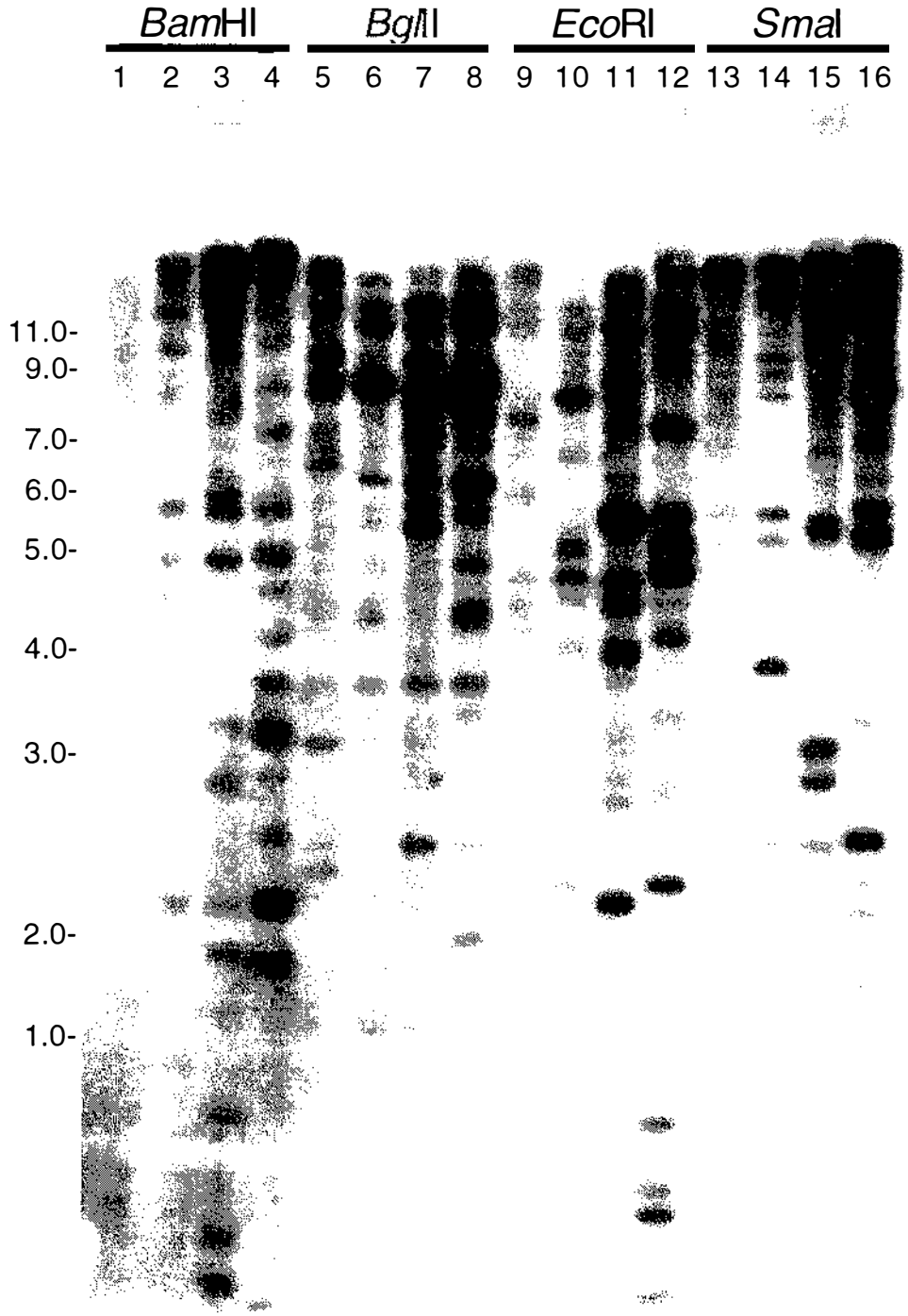



Figure 3.27 The 5S rRNA genes are dispersed in the *Epichloë* endophytes

0.8 µg of genomic DNA from Lp1A0 (lanes 1, 5, 9 and 13), Lp5 (lanes 2, 6, 10 and 14), E8 (lanes 3, 7, 11 and 15) and *Epichloë festucae* Frc7 (lanes 4, 8, 12 and 16) was digested with *Bam*HI, *Bgl*II, *Eco*RI and *Sma*I as indicated. These digests were fractionated on a 0.7% agarose gel and the resulting Southern blot probed with the 5S rRNA PCR product from primers 5SL - 5SR. Sizes of the 1 kb ladder size marker (Gibco BRL) are indicated in kb to the left.

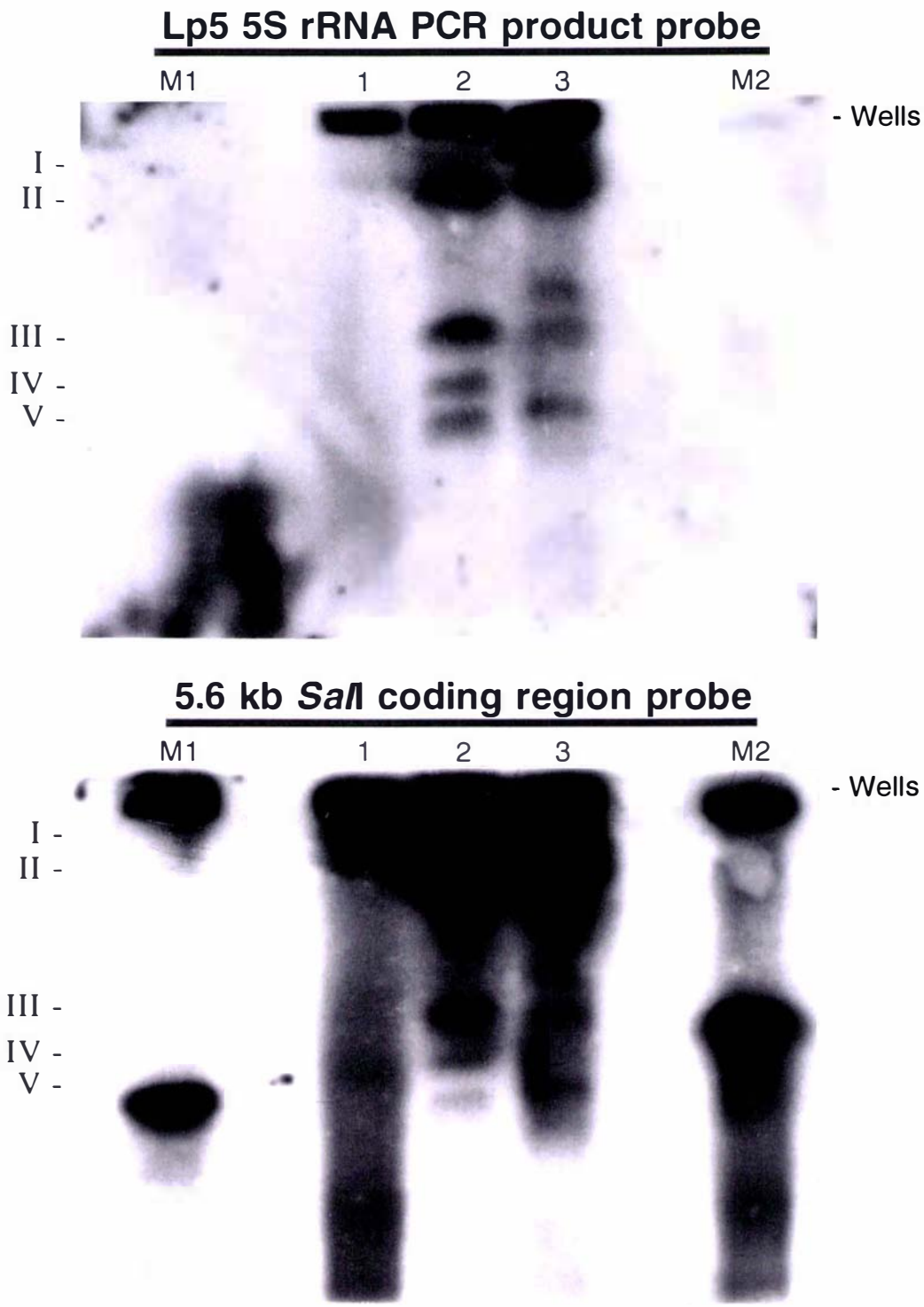



Figure 3.28 The 5S rRNA and *rrn* genes are located on the same chromosomal bands

The 1.0% chromosomal grade agarose CHEF gel Southern blot from Figure 3.18B was stripped and probed with the 5SL - 5SR PCR product from Lp5 as indicated. The blot from Figure 3.18B probed with the 5.6 kb *Sal*I coding region probe is shown for comparison, as indicated. Isolates are E8 (lane 1), Lp5 (lane 2) and Lp1A0 (lane 3). The chromosomal bands that hybridise in Lp1 are labelled from I to V. Lanes M1 and M2 are the *S. cerevisiae* and *S. pombe* chromosome size markers respectively.

# CHAPTER FOUR

# DISCUSSION

# 4.1 LENGTH HETEROGENEITY IS A RESULT OF UNEQUAL CROSSING OVER IN THE rDNA IGS

The results reported here reveal a high level of length heterogeneity in the ribosomal IGS region of the hybrid *Epichloë* endophyte isolate, Lp1. I have shown that this heterogeneity is intragenomic and is not the result of differences between cells. The length heterogeneity arises from copy number variation of a sub-repeat class, the 111/119 bp sub-repeats, that are located within the IGS. I propose that this copy number variation is a consequence of unequal crossing over occurring in the register of the 111/119 bp sub-repeats. Unequal crossing over, as discussed in Section 1.3.2, produces two reciprocal products and each contains a different number of repeats from the original molecules. Copy number variation of repeats is therefore believed to result from unequal crossing over between such repeats. IGS length heterogeneity as a consequence of copy number variation of sub-repeats has been reported in a number of types of organisms previously. Examples include vertebrates (Arnheim and Kuehn, 1979; Botchan *et al.*, 1977; Krystal and Arnheim, 1978; Wellauer *et al.*, 1976a), insects (Coen *et al.*, 1982b; Israelewski and Schmidt, 1982; Schäfer *et al.*, 1981; Tautz *et al.*, 1987; Wellauer and Dawid, 1978), plants (see Rogers and Bendich, 1987b), and fungi (Martin, 1990; Van Heerikhuizen *et al.*, 1985). However the extent of the length heterogeneity, the fact that it is clearly intragenomic and the sudden appearance of the heterogeneity from systems that do not display such heterogeneity (the two progenitor isolates) make Lp1 unusual.

## 4.1.1 UNEQUAL CROSSING OVER AS A MECHANISM FOR HOMOGENISATION OF THE rDNA

The unequal crossing over I have demonstrated within the IGS sub-repeats suggests that it plays a role in the concerted evolution of the whole rDNA. If misalignment of the 111/119 bp sub-repeats is concomitant with misalignment of entire rDNA units, then unequal crossing over will play a role in the concerted evolution of the Lp1 rDNA. There is no *a priori* reason to suspect that misalignment of the sub-repeats is occurring between directly equivalent rDNA units in the array. In fact misalignment of whole rDNA units may be more likely than misalignment of the sub-repeats, because the length of homology is much longer. Such unequal crossing over between whole rDNA units will produce copy number variation of rDNA units in the array (Dover *et al.*, 1993). Variation in the number of rDNA units between individuals has been demonstrated almost universally where it has been looked for (e.g. see Appels and Honeycutt, 1986; Flavell and Martini, 1982; Long and Dawid, 1980), which suggests that unequal crossing over in the register of whole rDNA units is a common phenomenon. Therefore I propose that unequal crossing over occurring between whole rDNA units is playing a role in the homogenisation of the rDNA in Lp1.

These data present an apparent paradox - the process that plays a role in the homogenisation of repeats (unequal crossing over) is responsible for generating a high level of heterogeneity in these repeats. This paradox is expected as the homogenisation process is working at the sequence level, while the misalignment that drives the sequence homogenisation is producing heterogeneity at the level of length (copy number variation of the sub-repeats). Unequal crossing over will tend to spread a particular repeat throughout the array, but as this repeat spreads, different copies acquire different numbers of sub-repeats strictly as a result of the process of spread (Dover, 1987; Williams and Strobeck, 1985).

### 4.1.2 UNEQUAL CROSSING OVER IN THE Lp1 rDNA IS MITOTIC

Although the distinction between mitotic and meiotic recombination is well appreciated, there has been debate on the relative roles that these different forms of recombination play in concerted evolution (Appel and Gordon, 1996; Campbell *et al.*, 1997; Crease and Lynch, 1991; Sang *et al.*, 1995; Shufran *et al.*, 1997). The unequal crossing over we have demonstrated here is strictly mitotic, as Lp1 is an asexual organism (M. Christensen, personal communication; Collett, 1994). There is evidence, aside from the lack of multigene family breakdown in asexual organisms, that mitotic recombination plays an important role in concerted evolution. Both unequal crossing over (Szostak and Wu, 1980) and gene conversion (Gangloff *et al.*, 1996; Muscarella and Vogt, 1993), the two principle mechanisms proposed to account for homogenisation, occur in mitosis in the rDNA. Interestingly, the rate of mitotic recombination is normally low in the genome, with the rDNA being an exception to this rule (Petes and Hill, 1988; Szostak and Wu, 1980).

Even in sexual organisms, much of the "work" of concerted evolution may occur during mitosis, with meiosis perhaps distributing homogenising arrays through the population, so that concerted evolution proceeds in a cohesive manner in sexual populations (Ohta and Dover, 1984; see Section 4.3.3). Alterations in rDNA number in both *Neurospora crassa* and *Drosophila melanogaster* are thought to occur premeiotically (Butler and Metzenberg, 1989; Hawley and Marcus, 1989) - during the mitotic divisions that lead to the production of the ascospores, rather than in meiosis itself (see Section 4.2). This is independent evidence for the importance of mitotic recombination in the regulation and homogenisation of the rDNA.

## 4.2 THE RATE AND NATURE OF TURNOVER IN THE IGS: INCONGRUENCES BETWEEN REPORTED RESULTS AND UNEQUAL CROSSING OVER MODEL OF HOMOGENISATION

Several of the results obtained do not fit neatly with the expectations generated from a simple model of unequal crossing as the primary homogenising mechanism. Detailed examination of these incongruences may help to give a better understanding of how homogenisation is achieved, and perhaps how it may be confounded.

### 4.2.1 RAPID TURNOVER OF REPEATS IN THE rDNA

The rate of turnover due to unequal crossing over in Lp1 is high. Two rounds of single-spore purification are sufficient to produce noticeable changes in the pattern of IGS lengths (Figure 3.2). Furthermore, this turnover is able to produce drastic changes in a relatively short space of time. Four of the hybrid laboratory cultures presented here, Lp1A, Lp1C, Lp1D and Lp1F, were derived from the initial isolate culture, Lp1, and were maintained as separate plate cultures for up to four years before the DNA used in this study was extracted. In this time they have evolved their own distinct banding profiles, arising from one initial profile. The independent LpTG-2 isolate, Lp2, also has its own distinct profile.

Rapid turnover of repeats in the rDNA is also shown by the spread of a deletion in the IGS which is independent of the 111/119 bp sub-repeats (Section 3.1.10). This deletion appears to be entirely lacking from Lp1 when it was first isolated (Figure 3.13), but has spread to relatively high levels in some laboratory cultures derived from it, and is absent in others. The deletion appears to have arisen from an intrachromosomal recombination event involving the 40 bp sub-repeats, and has resulted in the deletion of six of these sub-repeats. It does not appear to be the result of excision of a transposon or another mobile element.

The deletion is very likely to have occurred just once, and has spread through the rDNA repeats and Lp1 cultures since. Although the presence and absence of the deletion in various laboratory cultures can easily be explained by a sectoring phenomenon, with this sector then by chance being used to start some of the laboratory cultures, the relative stability of the deletion through single-sporing is evidence that the deleted IGS variant is present in a significant number of rDNA units in the genome. This suggests that the deletion has spread through many rDNA repeats in the time since it arose, presumably as a passenger of homogenisation. Thus a significant level of homogenisation can be achieved stochastically in a maximum of four years of mitotic growth. Indeed the discovery of multiple rDNA loci in Lp1 (Section 3.2) raises the question whether the deletion is restricted to a single locus, and perhaps has reached fixation in some cultures at that locus, given that homogenisation between loci appears to be much slower than that within a locus (Section 1.3.2).

### 4.2.2 CLUSTERING OF IGS LENGTH VARIANTS

The changes in IGS banding pattern through the two rounds of single-spore purification have an unexpected feature. Several of the bands that either appear or disappear through the single-sporing are strongly hybridising (Figure 3.2), and therefore must represent a number of copies of that particular IGS length. Unequal crossing over between rDNA units will result in the stochastic loss or gain of IGS lengths. Therefore loss of a strongly hybridising band is likely to represent the loss of a block of rDNA units all with the same IGS length. This implies that rDNA units with the same IGS lengths are clustered together. This is plausible for both an unequal

crossing over or a gene conversion scenario where the degree of misalignment is small (Kimura and Ohta, 1979). Szostak and Wu (1980) found the degree of misalignment in yeast rDNA mitotic unequal crossing over was six to eight units, and interestingly this corresponds well with that found in *D. melanogaster* 5S rRNA gene unequal crossing over (Samson and Wegnez, 1988). Similarly there is evidence to suggest that adjacent and alternate repeats tend to be similar, perhaps indicating a localised mode-of-action of the homogenisation mechanism (Dvorák *et al.*, 1987; Lassner and Dvorák, 1986). This evidence includes potential clustering of length variants in an array, but there has been considerable debate in the literature as to whether clustering occurs (Barker *et al.*, 1988; Beckingham, 1981; Coen and Dover, 1983b; Copenhaver and Pikaard, 1996b; Dvorák and Appels, 1986; Hawley and Tartof, 1983; Pellegrini *et al.*, 1977; Renkawitz-Pohl *et al.*, 1981; Rogers and Bendich, 1987a; Sharp *et al.*, 1983; Tulin *et al.*, 1997; Wellauer *et al.*, 1976b; Williams *et al.*, 1990). Evidence for both states may indicate differences between species or a moderate level of clustering.

I also show the appearance of strongly hybridising bands. This does not necessarily imply that rDNA units with the same IGS length are clustered, but the loss of strongly-hybridising bands discussed above suggests they are. Given the vast range of possible IGS lengths, it is also likely that most IGS length variants result from a single event and are then amplified, presumably by the homogenisation mechanism. If so, the amplification must be very rapid, and several sequential misalignment events would be required in the amount of mitotic growth it takes to generate a fungal disk from a single-spore. The process also appears to be directional in the sense that only a very few new length variants are altered, but those that do change by a significant number of copies. These results are highly unexpected given the documented rates of unequal crossing over in the rDNA and its stochastic nature.

### 4.2.3 RESTRICTION OF IGS LENGTHS

The IGS lengths we see do not represent a clean ladder of bands at ~ 115 bp intervals as one might expect if the length varies as a result of the 111/119 bp sub-repeats. In a sense this is anticipated, as there are such an enormous number of IGS lengths possible (the largest IGS lengths observed in this study would contain over 150 sub-repeat copies) that statistically one would not expect to find all lengths represented. Even so, there are only a very limited number of lengths observed (Figure 3.2; many of the cultures contain only 8 - 16 predominant hybridising bands). Lp1A in particular shows a very skewed range of IGS lengths, falling almost entirely in either the high or low molecular weight part of the range.

Restriction of IGS lengths to a small number of the total possible set is presumably the result of a homogenisation process that (stochastically) amplifies the number of these particular IGS lengths at the expense of others. This homogenisation process is unlikely to be unequal crossing over, as unequal crossing over tends to increase IGS length variation by generating new IGS lengths whilst homogenising existing ones, as long as there is misalignment of the sub-repeats. The net

outcome of this process is maintenance of length variation. It is possible that the majority of unequal cross over events are equal with respect to the sub-repeats (e.g. they may occur outside of the sub-repeats), but if this was the case the rate of unequal crossing over must be much higher than that suggested by looking at turnover of IGS length variants, which seems unlikely.

The IGS length variants also seem restricted in the minimum size they can attain. The size of the "original" IGS in Lp1 can be assumed to be around 4 kb, as both of the Lp1 progenitors have IGS lengths around 4 kb. This leads to an interesting conclusion - the length heterogeneity in Lp1 is almost exclusively an increase in length. Yet the reciprocal nature of unequal crossing over dictates that for every larger product formed, a smaller product must also be formed. The smallest IGS length is 3.5 kb, which would contain nine 111/119 bp sub-repeats - enough for more to be lost.

### 4.2.4 NON-RANDOM ARRANGEMENT OF THE IGS SUB-REPEAT VARIANTS

In repeat arrays that are shaped by the forces of unequal crossing over, variants that arise and become eliminated from the array do so by being moved to the edges of the array, a phenomenon known as terminal exclusion (Barker *et al.*, 1988; Dover *et al.*, 1993; Dvorák *et al.*, 1987; Kamnert *et al.*, 1998; Mian and Dover, 1990; Tulin *et al.*, 1997). This occurs because the cross over point is assumed to be random. The MVR-PCR analysis we performed gives the order of the 111/119 bp sub-repeats in the clone, and these results show that terminal exclusion is not occurring for this IGS (Section 3.1.8). The results from genomic DNA corroborate this.

The final incongruence between the results presented here on the IGS length variation and an unequal crossing over model for homogenisation is the behaviour of the 111/119 bp sub-repeats in genomic DNA. The unequal crossing over model of homogenisation has assumed a random cross over point (Linares *et al.*, 1994; Smith, 1976). It follows that with crossing over occurring randomly in the sub-repeat array, both ends of the array should behave the same. However the results from Section 3.1.8 show that while the order of the two types of sub-repeat on the right hand end of the array is highly conserved, the order on the left hand end is highly variable. Clearly the two ends of the array are behaving differently.

## 4.3 IMPLICATIONS FOR THE HOMOGENISATION MECHANISM

Some of the incongruences outlined in the above section can be explained by modifications of the unequal crossing over model for homogenisation, but for the others alternatives to the unequal crossing over model of homogenisation should be considered.

### 4.3.1 SELECTION AGAINST SMALL SPACERS

The lack of IGS length variants in Lp1 smaller than around 3.5 kb suggests a form of "selection" against short spacers. I propose some sort of homology interaction. It seems likely that the recombination equipment is limiting (Gottlieb *et al.*, 1989; Kanaar and Hoeijmakers, 1998; Loidl and Nairz, 1997; Wu and Lichten, 1995), leading to competition between rDNA units for this equipment. (Jinks-Robertson *et al.*, 1993; Yuan and Keil, 1990) found that recombination in yeast between non-tandem duplications requires at least 250 bp of homology, and the rate increases linearly up to 1 kb of homology, after which it plateaus off. Nine 111/119 bp sub-repeats represent approximately 1 kb, therefore efficient recombination might only be achieved by rDNA units with at least this number of sub-repeats. It should be noted that homologous recombination in yeast is particularly efficient, so other organisms may require more homology for efficient recombination. Sugawara and Haber (1992) on the other hand found that only about 80 bp of homology were required in a similar system. The reason for the discrepancy is not clear. Spread of a repeat through an array by unequal crossing over requires that repeat to participate in the unequal crossing over (Smith, 1973), so any repeats not participating in unequal crossing over, such as those with short 111/119 bp sub-repeat arrays, will be eliminated from the array as other repeats spread.

An alternative to this model is gene conversion. It has been proposed that gene conversion can change the copy number in a repeat array (Fogel *et al.*, 1984; Gangloff *et al.*, 1996; Welch *et al.*, 1990), and this occurs without the reciprocal change in copy number that is required for unequal crossing over. This model would then posit that gene conversion is responsible for the length heterogeneity seen in the Lp1 IGS, and that this gene conversion is biased towards a gain of sub-repeats. This is discussed in more detail in Section 4.4.

### 4.3.2 BIASED INITIATION OF CROSSING OVER IN THE SUB-REPEATS

The different behaviours of the ends of the 111/119 bp sub-repeat array in genomic DNA suggests that the cross over point is not random, and instead we should consider specific initiation of unequal crossing over. There does, in fact, seem to be little biological justification for assuming a random cross over point. Interestingly, a similar phenomenon has been found in the human M32 minisatellite array (Jeffreys *et al.*, 1990) and the primate U2 snRNA tandem array (Pavelitz *et al.*, 1995). These were explained by gene conversion events initiating preferentially at one side of the array (Jeffreys *et al.*, 1994), and I propose a similar model for the results presented here.

Many advances have been made in recent years in understanding the biochemistry of recombination. In the best-studied system, the Chi system in prokaryotes (reviewed in Eggleston and West, 1996), a double strand break (DSB) is made and the RecBCD complex unwinds and degrades the DNA until it encounters a Chi sequence in the correct orientation on the 3' to 5'

strand (Figure 4.1A). The DNA is then cleaved near the Chi sequence, and further unwinding generates a single strand tail with Chi at its 3' end which initiates pairing and strand exchange with the help of RecA. Thus Chi has both orientation dependence and directionality. Holliday junctions are then formed, and the heteroduplex DNA is extended by branch migration via the action of RuvAB. Finally resolution of the Holliday junctions to generate the recombinant molecules is catalysed by RuvC, which preferentially nicks the DNA at a short target sequence (Shah *et al.*, 1994). Thus there are potentially three sites that mediate recombination - initiation of DSB, a Chi-like sequence to trigger strand exchange, and a signal to resolve the Holliday junctions.

Biochemical understanding of eukaryote recombination lags behind that of prokaryotes, but many features seem to be conserved (Cox. 1997). A broadly similar model for recombination in the Lp1 rDNA involving these three sites of recombination mediation would lead to non-random cross over points (Figure 4.1B). If a DSB was initiated to the left of the 111/119 bp sub-repeats (shown in the 40 bp sub-repeats in Figure 4.1B; see below) and there was a Chi-like sequence at or to the left of the sub-repeats, then Holliday junction resolution favoured by a sequence within the sub-repeats would bias the point of cross over to the left-hand side of the sub-repeat array (Christiansen *et al.*, 1987; Evans and Kolodner, 1988; Sekiguchi *et al.*, 1996). The extent of branch migration would determine how far to the right the cross over point occurred. Bias of cross over to the left-hand side of the sub-repeat array (in the orientation shown in Figure 3.10) would conserve the sub-repeat order at the right-hand end of the array while tending to rearrange the order at the left-hand end, unless misalignments occurred with the right-hand end of one of the sub-repeat arrays. This is consistent with the observed results. Whilst Smith (1994) has suggested that a pathway similar to Chi may operate in eukaryotes, there are at least some recombination pathways that seem to be unique to eukaryotes (Haber, 1999).

Linares *et al.* (1994) in their examination of the unusual sub-repeat organisation in the *D. melanogaster* rDNA IGS implicated a smaller, more poorly maintained sub-repeat array containing simple sequence motifs in the initiation of gene conversion that shapes the larger sub-repeat array. Other studies have also implicated sequence simplicity in homogenisation (Barker *et al.*, 1988; Eickbush and Burke, 1986; Lassner *et al.*, 1987; Liao *et al.*, 1997). It is interesting in this regard that the Lp1 IGS also contains a smaller, poorly maintained sub-repeat array with simple sequence motifs (the 40 bp sub-repeats) at the left-hand side of the 111/119 bp sub-repeat array. Sub-repeat arrays with simple sequence motifs may act as recognition sites for the initiation of recombination (unequal crossing over and/or gene conversion), perhaps as the site of DSB. There is evidence for recombination-initiating DSB in the rDNA (Høgset and Øyen, 1984; Rasooly and Robbins, 1991; Rockmill *et al.*, 1995), and the discovery of fragile sites in the rDNA of *N. crassa* (Butler, 1992) may be a consequence of such an activity. Also of note, the 111 bp sub-repeat contains a sequence (AGTGGTGG; Figure 3.8B shows the reverse complement of this sequence) that shows a high degree of similarity to the Chi sequence (GCTGGTGG). This is in the orientation that would stimulate recombination if the DSB

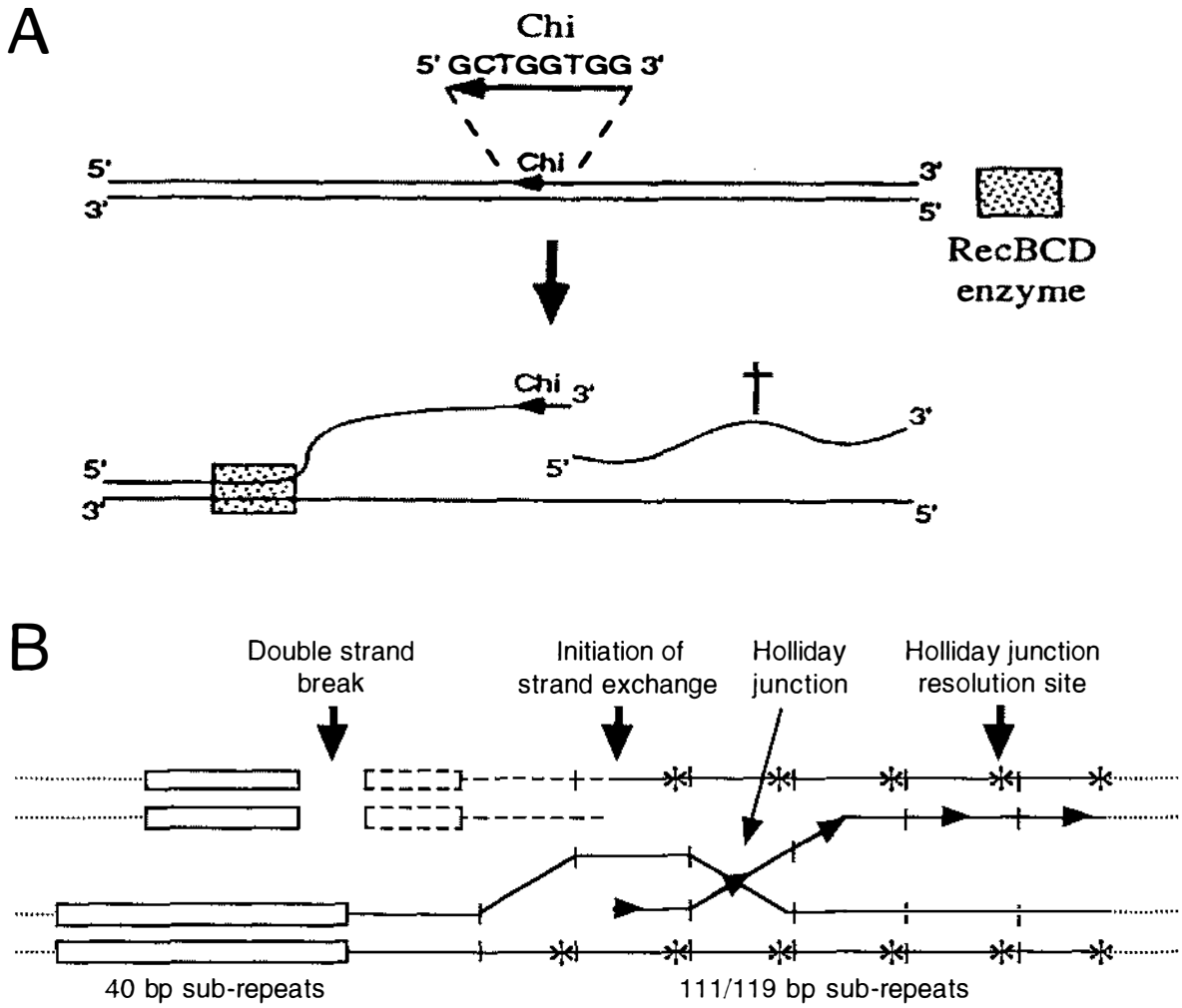



Figure 4.1 Model of biased initiation of crossing over based on the Chi system in prokaryotes

A Diagram indicating the orientation dependence of Chi. If the RecBCD enzyme encounters a Chi sequence orientated as shown, there is a nick made at Chi. The nuclease activity of the RecBCD enzyme is attenuated, but it continues to unwind the DNA past Chi. The single stranded DNA marked with † is degraded. The single stranded DNA tail with Chi at the 3' end then initiates the rest of the recombination pathway. Diagram taken from Smith (1994).

B Stylised diagram indicating how the cross over point might be biased to one side of the 111/119 bp sub-repeat array in the Lpl rDNA IGS. The hypothetical model begins with a double strand break in the 40 bp sub-repeats (boxed in grey). The DNA is degraded (broken lines) until an initiator of strand exchange, analogous to Chi, is reached (indicated by arrowheads in the 111/119 bp sub-repeats, which are delineated by short, vertical lines). The strand exchange results in the formation of a Holliday junction, which can be resolved by cleavage at specific sites (shown as asterisks in the 111/119 bp sub-repeats) following branch migration. Cleavage of the outer strands will result in cross over products. Each line represents one strand of the DNA helix; the two interacting helices could be sister chromatids. Flanking sequence is indicated by dotted lines. Diagram not drawn to scale.

initiation point was to the left-hand side of the sub-repeat array (Figure 4.1B). This sequence contains the two paired guanosines that Hibner *et al.* (1991) found to be present in many recombination-stimulating sequences.

The restriction of spacer lengths to a small percentage of those possible does not fit well with unequal crossing over as the homogenisation mechanism. Once again this may be accounted for by bias in the initiation of crossing over. Williams and Strobeck (1985) showed that with a very simplified model of IGS length variation generated by unequal crossing over, the number of length variants in an array could vary dramatically, from very many to very few, strictly as a result of the probabilities of unequal versus equal crossing over in the register of both whole rDNA units and IGS sub-repeats. As noted above, a cross over event occurring outside the IGS sub-repeats is seen as an equal cross over with respect to these sub-repeats and can thus homogenise length variants. However this would imply that the rate of unequal crossing over in the rDNA is much higher than that seen within the sub-repeats. The presence of multiple rDNA loci in Lp1 (Section 3.3.2) may indicate that different groups of IGS length variants are restricted to different rDNA arrays. An alternative scheme is that gene conversion is operating at a similar rate to unequal crossing over, producing an equilibrium between the generation and elimination of IGS length variants.

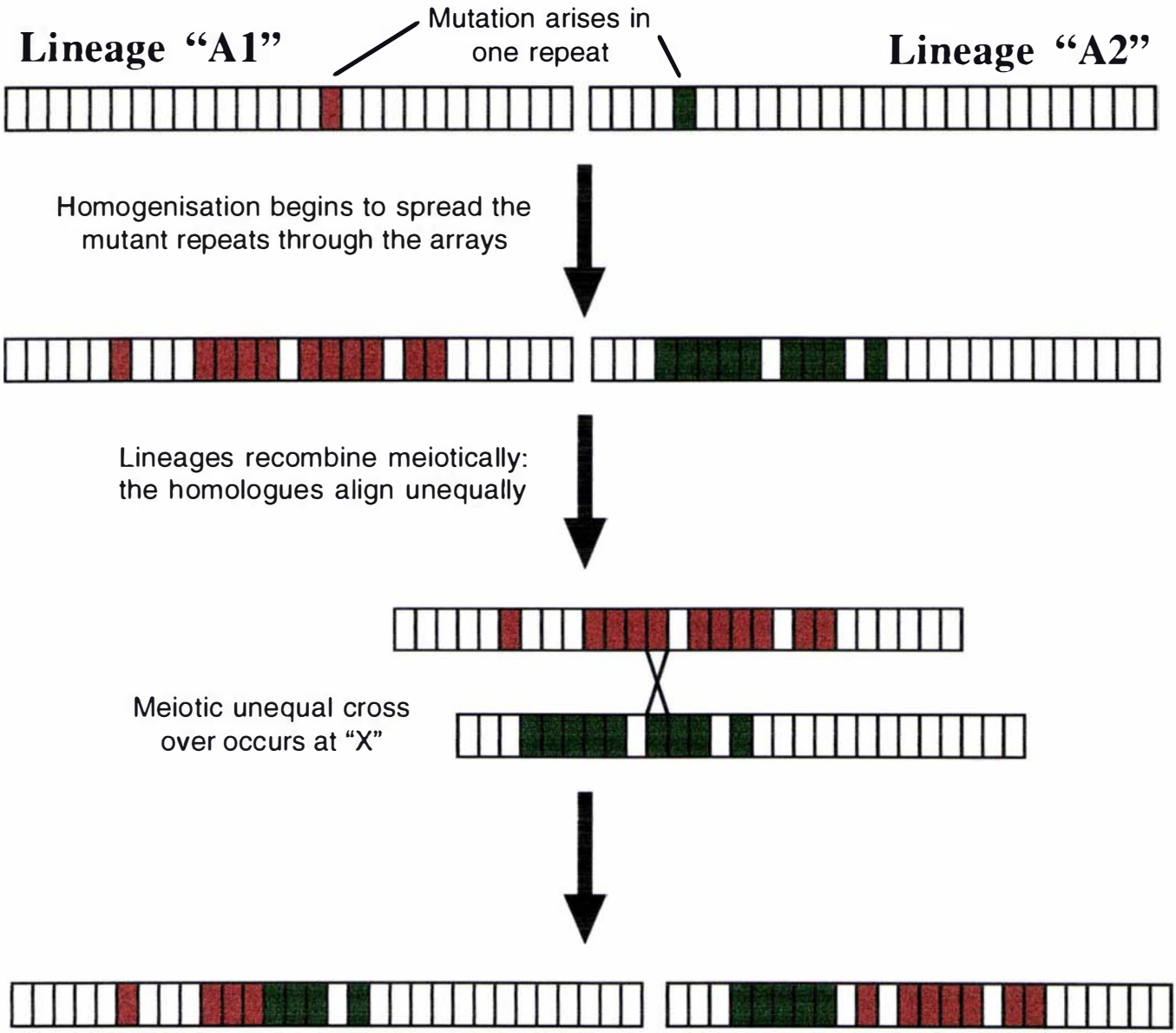
### 4.3.3 RAPID HOMOGENISATION

The generation of distinct IGS length variant patterns in different laboratory cultures of Lp1 in a handful of years and the rapid spread of an IGS deletion variant in the same time argue for a rapid turnover of rDNA repeats. Indeed the large increase in the proportion of deleted versus non-deleted spacer in one round of single-sporing (Section 3.1.10) argues that a significant level of spread of this variant occurs in the amount of mitotic growth it takes to generate a colony from a Single-spore. Coen *et al.* (1982a) estimated the rate of homogenisation, based on unequal crossing over, to be in the range of fixation of a variant in an array in 1-10 x $10^6$ generations for *Drosophila*. However the authors suggest that the real rate must be somewhat faster, as this is slower than the proposed speciation rate of *Drosophila* (Dover *et al.*, 1982). Although it is not easy to relate generation times of a sexually-reproducing animal to the growth of an asexual fungus, the results presented here suggest homogenisation in Lp1 at least is much faster than this, and that fixation of a variant should be observable in biological time-frames. It is worth noting that other authors have suggested that homogenisation is "rapid" (see especially Fukuoka *et al.*, 1994; Liao *et al.*, 1997), although in many cases the time scale they refer to is not clear (Dvorák and Appels, 1982; Jakubczak *et al.*, 1992; Kamnert *et al.*, 1998; Strachan *et al.*, 1985; Williams *et al.*, 1985).

Is Lp1 exceptional in its rate of homogenisation? Although Lp1 seems anomalous in its extraordinary level of IGS length variation, I do not believe that this translates into an exceptionally high rate of homogenisation. As will be discussed in Section 4.6, generation of the

IGS length heterogeneity appears to be a consequence of the hybridisation. However I believe that turnover of IGS length variants is indicative of a "normal" rate of homogenisation, rather than a speeding up of processes that are already occurring in the progenitors, and instead the length heterogeneity is an indication of genomic perturbation that has occurred as a result of the hybridisation. That is not to say that there is no change in rate in the hybrid, but I believe any change will have been relatively small and cannot account for the discrepancy between the rates in the literature and that implied by my results.

Why such a discrepancy? The only reported attempt to measure the rate of homogenisation experimentally was by Coen *et al.* (1982a) and involved a sexually-reproducing species. I propose that meiotic recombination actually functions to reduce the rate of homogenisation that occurs mitotically. In general meiotic recombination introduces variants from one homologue onto another. This has been exploited by sexually-reproducing organism to spread favourable mutations to many individuals in a population so the favourable mutation can quickly reach fixation. In the context of a multigene family, meiotic recombination also acts to spread variants, but the short-term outcome will be different. Consider two mitotically-reproducing lineages (Figure 4.2). In one a variant, A1, arises and is spread to a subset of repeats, and the same happens in the other lineage, but with variant A2. Now imagine that these two lineages meiotically recombine. The progeny will now contain A, A1 and A2 in the repeat array. A1 and A2's spread to fixation in the array is retarded, although their spread through the population is increased. A's spread to fixation in the array may be increased, retarded or the same depending on the chance position of the cross over point. Experimentally this has been recognised to some extent (Appel and Gordon, 1996; Crease, 1995; Shufran *et al.*, 1997), but the implication (i.e. that measurement of the rate of homogenisation should be done in an asexual system) has apparently not been made. Therefore I propose that an accurate measurement of the rate of homogenisation will be possible in an asexual system, and the rate of homogenisation is rapid enough to be monitored in "real time". Unfortunately the *Epichloë* endophytes are not an ideal system to undertake this measurement. The presumed slow rate of interchromosomal homogenisation (Section 1.3.2) would make observation of the spread of a variant throughout the rDNA very complicated in a system such as these endophytes where there are multiple rDNA arrays (Section 3.2). If the rate of homogenisation is very rapid then various results will need to be re-examined. For instance, the cohesive spread of an array though a population (Ohta and Dover, 1984) which relies on the rate of randomisation of chromosomes by sex being faster than homogenisation (Dover, 1986), would not be expected. There is some evidence for this already (Crease and Lynch, 1991; Schlötterer and Tautz, 1994). On the other hand, persistence of polymorphisms within an array would not be expected.




## Figure 4.2 Meiotic recombination disrupts homogenisation

Diagrammatic representation of the ability of meiotic recombination to retard the spread of variants that are being homogenised in the array. Boxes are representative of rDNA units, with wild type A rDNA units shown as open boxes, and mutant variants A1 and A2 shown as red and green boxes respectively. The figure shows a variant spreading in each of two lineages which are initially evolving separately. These lineages then combine through a sexual recombination event. This retards the spread of the variants, and may accelerate or retard the homogenisation of the original A unit.

## 4.4 GENE CONVERSION AS A HOMOGENISATION MECHANISM

The MVR-PCR analysis in Section 3.1.8 did not give any clear evidence for terminal repeat exclusion of one of the 111/119 bp sub-repeat variants, as is predicted for homogenisation by unequal crossing over. Two alternative, but not mutually-exclusive, explanations can be made for this lack of terminal exclusion. Either the 111 and 119 bp sub-repeats are both required for some role in the IGS, especially its concerted evolution, or other mechanisms alongside unequal crossing over are also shaping this sub-repeat array. The other oft-cited mechanism proposed for homogenisation is gene conversion, therefore gene conversion could be responsible for the lack of terminal repeat exclusion. Several authors have proposed a combination of mechanisms to account for homogenisation (Appels and Honeycutt, 1986; Birky and Skavaril, 1976; Coen *et al.*, 1982b; Dover, 1987; Dover *et al.*, 1982; Jinks-Robertson and Petes, 1993; Linares *et al.*, 1994; Tautz *et al.*, 1987; Williams *et al.*, 1989) However, it should be noted that some of these authors have proposed that unequal crossing over results in homogenisation within an array, while gene conversion is responsible for exchanges between homologous and especially non-homologous chromosomes. An almost-opposite proposal has been made for homogenisation of the rDNA in plants (Zhang and Sang, 1999; see also Hadjiolov 1984). Detailed studies of the late chorion multigene family in *Bombyx mori* have presented convincing evidence for homogenisation of the repeats occurring through a "patchwork" of gene conversion events (e.g. Eickbush and Burke, 1985).

In Section 4.3 I have also suggested gene conversion as an alternative explanation for some of the incongruences between my results and an unequal crossing over model of homogenisation. An interplay of unequal crossing over and gene conversion is attractive because both mechanisms have been experimentally demonstrated, and they can each be used to explain different patterns observed in multigene families. However experimental evidence for their action in homogenisation is difficult because of their mechanistic similarity and the difficulty in separating the two mechanisms to test predictions.

### 4.4.1 A RAPID HOMOGENISATION-LIKE PROCESS INDUCED BY DOUBLE STRAND BREAKS

A set of very interesting experiments was done by Muscarella and Vogt (1993). They were looking at the expression of the endonuclease I-*Ppo*I (used in this study) that cleaves the 25S *rrn* gene in *Saccharomyces cerevisiae*. When they expressed this endonuclease, very few colonies were recovered, presumably because the DSBs created in the rDNA were lethal. Several of the surviving colonies had mutations in the 25S *rrn* gene at the recognition sequence of the endonuclease, rendering the rDNA resistant to the activity of the endonuclease. There were three classes of mutation. Interestingly, each colony recovered had only a single class of mutation present in all the rDNA units. They propose that a mutation occurs in a single rDNA unit, and

this then spreads to all the other copies. The spread is believed to occur by gene conversion, which is initiated by the DSBs created by the endonuclease in rDNA units without the mutation. The only colonies that survive have the mutation present in all the units, as a DSB in mitosis is lethal, and therefore all the rDNA units must be converted before the first mitotic division. This is obviously a "concerted" event, with the spread of a mutation through all the rDNA unit copies being driven by the induced DSBs, and their resultant lethality, at a very fast rate.

It is not known whether this is relevant to "normal" homogenisation of the rDNA. That the DSBs induced by the endonuclease stimulate gene conversion has been established - this is how "intron homing" of group I introns encoding an endonuclease occurs (Muscarella and Vogt, 1989). The rarity of colonies that become "resistant" to the action of the endonuclease may reflect the probability of a mutation arising in the recognition site. It may also reflect the probability of the homogenisation machinery being able to spread the mutation to all rDNA units before the first mitotic division occurs, as the recombination machinery is believed to be limiting (Gottlieb *et al.*, 1989; Kanaar and Hoeijmakers 1998; Loidl and Nairz 1997), or it may reflect a combination of these two factors. Alternatively this spread may be the consequence of an action separate from homogenisation that occurs at a very low rate, but which is unmasked by the abnormal selection that generation of DSBs in the rDNA provokes.

The role played by gene conversion in homogenisation of the rDNA in Lp1 could in principle extend to the generation of copy number variation of the IGS sub-repeats, and hence the IGS length heterogeneity. Gene conversion is not restricted to conversion of base pair mismatches - it can also convert insertions and deletions (Whitehouse, 1982). In terms of tandemly-repeated sequences this can result in alteration of copy number if the repeats align with some repeats "looped out" of the aligning molecules. A gene conversion event incorporating such looped out sections will result in a change of copy number, either an increase or decrease depending on which DNA strands act as donor and recipient (Fogel *et al.*, 1984). Experimental evidence for this phenomenon has been obtained (Gangloff *et al.*, 1996; Welch *et al.*, 1990), but the initial proposal of Fogel *et al.* (1984) has attracted little attention. Indeed Gangloff *et al.* (1996) seem to propose a cross over event occurring alongside conversion to explain alteration in repeat number, and they suggest that this mechanism is solely responsible for homogenising the rDNA. Under such a scheme gene conversion could explain the IGS length heterogeneity in Lp1, removing the need to invoke unequal crossing over action in the rDNA at all. Whether gene conversion can generate the extensive copy number variation that is observed in the Lp1 IGS remains to be determined, but it presumably would involve a strong bias towards replacement of the non-looped out section with the looped out section, thus increasing copy number. Copy number variation of whole rDNA units (see Section 4.1.1) has been described almost exclusively in terms of unequal crossing over, and it remains to be seen whether gene conversion can act on sections of DNA as large as multiple rDNA units.

### 4.4.2 POTENTIAL ROLE OF DOUBLE STRAND BREAKS IN rDNA HOMOGENISATION

Clearly both unequal crossing over and gene conversion can be invoked to explain various aspects of the results presented in this study. The results of Muscarella and Vogt (1993) discussed above suggest that the rate of gene conversion, and therefore perhaps homogenisation, could be finely controlled by regulating the number of DSBs that occur in the rDNA. However this is not restricted to gene conversion - crossing over is stimulated by DSBs too (Szostak *et al.*, 1983). Therefore gene conversion and unequal crossing over may be different outcomes of the same initiating event. There is some evidence for the induction of DSBs in the rDNA (Butler, 1992; Høgset and Øyen, 1984; Rasooly and Robbins, 1991; Rockmill *et al.*, 1995).

Interestingly, Butler and Metzenberg (1989) found that the frequent changes in the number of rDNA units in *N. crassa* occurs premeiotically, that is in the mitotic cell divisions that lead to the production of meiotic ascospores. The defence mechanisms against repeats found in *N. crassa* (repeat-induced point mutations and methylation induced premeiotically) are both found to occur premeiotically. Also, it was found that magnification of rDNA units as a result of *bobbed* mutations (see Section 4.5.1) in *D. melanogaster* occurs premeiotically (Hawley and Marcus, 1989), as do rearrangements in the human M32 minisatellite array (Jeffreys *et al.*, 1990). Thus it appears that this premeiotic part of the cell cycle may be a critical time for regulation of rDNA unit number in many organisms. The premeiotic stage in fungi involves cell and nuclear divisions which follow nuclear "pickup", but precede nuclear fusion and the subsequent meiosis. It is possible that the majority of these cells/nuclei are dispensable. If so such a stage is ideal for any process that may result in a high level of cell mortality, such as induction of a high level of DSBs (or inactivation of potentially foreign DNA), a form of cell-line selection. I believe it is not unreasonable to suppose that if the cell did use DSBs to regulate homogenisation, it would restrict the usage to developmental stages where potentially lethal effects could be attenuated. Although the premeiotic stage may be critical, such events can also occur at other stages of the life cycle (Pukkila and Skrzynia 1993). Whether an equivalent event to the premeiotic period is maintained in asexual organisms is not known.

## 4.5 EXTRACHROMOSOMAL rDNA CIRCLE AMPLIFICATION AS A MECHANISM FOR HOMOGENISATION

The rapid accumulation of multiple copies of an IGS length variant that are clustered together is not explained by unequal crossing over (Section 4.2.2). Even if unequal crossing over is extremely rapid, it is difficult to understand why only one length variant gets amplified, rather than concomitant amplification of a number of length variants. Gene conversion does not seem adequate to explain this phenomenon either. Instead there seems to be a process which causes amplification of a particular length variant such that the amplified copies are clustered together in the array. Reeder *et al.* (1976) and Wellauer *et al.* (1976b) suggested that the transfer of a cluster

of a single spacer length class from one homologue to the other in *X. laevis* was a consequence of extrachromosomal amplification of rDNA units and their insertion into the cluster. This mechanism can explain both the amplification of multiple copies of a length variant, and the fact that they are clustered in the array. Therefore I propose that extrachromosomal circle amplification of rDNA units followed by reinsertion into the array is another mechanism of homogenisation.

### 4.5.1 PREVIOUS MODELS OF HOMOGENISATION

Before looking at extrachromosomal rDNA circle (ERC) amplification as a mechanism for homogenisation, I will first consider the history of mechanisms proposed to account for concerted evolution. Theoretical and experimental considerations of the maintenance of repeated sequences led to two broad classes of homogenisation mechanism being proposed (Brown and Sugimoto, 1973; Edelman and Gally, 1970). Gradual turnover mechanisms such as unequal crossing over and gene conversion represent one broad class. The mechanisms of master-slave correction and saltatory replication were proposed for the other class. Master-slave correction proposed that there is a "master" gene and a large number of "slave" genes. In each generation all the slave genes are periodically converted into the master gene sequence, thus achieving homogeneity (Callan and Lloyd, 1960). Saltatory replication proposed that in a multigene family some genes would be amplified and reinserted back into the array, replacing many of the previous genes. Cycles of this coupled with selection and/or drift would lead to homogeneity (Britten and Kohne, 1968). The master-slave model was discounted because of difficulties in understanding how the master gene could be distinguished, but primarily because the level of variation found within a multigene array was much higher than that predicted by this model.

Saltatory replication and unequal crossing over were alternative hypotheses put forward to explain magnification in *bobbed* mutants in *D. melanogaster*. *Bobbed* mutations present as phenotypes that are caused by deletion of a large number of rDNA units from the array (Hawley and Marcus, 1989). Magnification is a process whereby these deficiencies are corrected in the germ line cells of *bobbed* males, such that some of their progeny inherit an increased number of rDNA units and are consequently wild-type. The increase in rDNA copy number naturally led to the idea that the mechanism of magnification would be tied in with the homogenisation of the rDNA. An instability in the initial generation of *bobbed* revertants coupled with the discovery of a transient increase in the amount of rDNA (subsequently shown to be ERCs) in the germ line cells of males undergoing magnification strongly pointed to a kind of saltatory replication being responsible for magnification (Graziani *et al.*, 1977; Ritossa *et al.*, 1973). However Tartof (1973) showed that magnification was prevented in *D. melanogaster* strains that contained a ring X chromosome. This was a "body-blow" for saltatory replication, as the dicentric product of an unequal cross over between sister chromatids with a ring X chromosome will be lost, but a ring X chromosome should not affect extrachromosomal amplification and integration. Unequal crossing over, especially between sister chromatids, then became the dominant paradigm for

homogenisation. It was joined by the mechanistically-related gene conversion, from a theoretical perspective because of its explicitly homogenising action, and also because it was able to explain interlocus homogenisation that unequal crossing over could not. Since then evidence has been presented questioning the role of unequal crossing over in magnification (Terracol, 1987), as well as explaining it in terms of unequal exchange between the X and Y chromosomes, producing an X-$Y^L$ chromosome (Frankham, 1988; Gillings *et al.*, 1987). It is testament to the enormous complexity of this process that even now the molecular details are unclear (Endow and Atwood, 1988; Hawley and Marcus, 1989).

Magnification appears to be associated with what seems to be an exquisite control of rDNA copy number, and therefore the processes involved in this form of regulation may not be directly tied in with the homogenisation mechanism (s). This regulatory control of copy number has been most elegantly demonstrated in *N. crassa*, where a process that resembles magnification also occurs (Russell and Rodland, 1986). The authors demonstrated that strains with deletions of rDNA units recovered to wild-type rDNA unit levels in a few generations. Furthermore strains that had increased numbers of rDNA units also returned to wild-type levels, albeit more slowly (Rodland and Russell, 1982). These results suggest that regulation of rDNA copy numbers occurs in two directions - ensuring both a minimum and maximum size for the array. Deletions of rDNA units also revert back to wild-type level in about 150 generation in *S. cerevisiae* (Kaback and Halvorson, 1977). Brewer *et al.* (1992) found that when the entire rDNA in *S. cerevisiae* was rendered non-transcriptional, 80% of the rDNA units were lost. It remains to be determined if this is a general phenomenon, but if so it would be interesting to test whether this loss of rDNA units is related to the stimulation of recombination by transcription (Stewart and Roeder, 1989).

Since the late seventies many investigators have looked for evidence of unequal crossing over in the rDNA, primarily looking at IGS length variation as in this study, and in most cases have been successful. In hindsight this is no surprise. Recombination is more-or-less a genome-wide phenomenon, and unequal exchanges are common occurrences wherever there are repetitive elements. This holds true for both meiotic and mitotic recombination - indeed there is actually a suppression of meiotic recombination between homologues in the rDNA array (Petes and Botstein, 1977; Russell *et al.*, 1988). Of course, finding unequal crossing over in the rDNA does not necessarily mean that it is responsible for the homogenisation of the rDNA. Where investigators have looked for patterns of repeats that are predicted by unequal crossing over, they have been less successful (Section 3.1.8; Linares *et al.*, 1994).

### 4.5.2 EXTRACHROMOSOMAL rDNA CIRCLE AMPLIFICATION MODEL OF HOMOGENISATION

The above considerations, in conjunction with the results I discussed in Section 4.2.2, have led me to propose that ERCs may play a role in the homogenisation of the rDNA in a manner reminiscent of the saltatory replication model. The general model proposes that an ERC containing one or more rDNA units is formed, with an intrachromatid exchange being the most likely mechanism. This ERC then undergoes rolling circle replication (Gilbert and Dressler, 1968; Hourcade *et al.*, 1973) to produce many copies of itself, and these copies reinsert back into the rDNA array (Figure 4.3). Buongiorno-Nardelli *et al.* (1972) proposed a form of saltatory replication involving extrachromosomal ribosomal DNA to account for concerted evolution of the rDNA. Details of their model proved to be false (see MacGregor 1982), but as far as I know the broad scheme of their model remains valid.

This ERC amplification model would require the rDNA unit to contain an autonomously-replicating sequence (ARS). Replication origins are commonly found in the rDNA (Hadjiolov, 1984). No sequence that matches the ARS in *S. cerevisiae* (Broach *et al.*, 1982), which can also act as an ARS in other fungi (e.g.Cannon *et al.*, 1990), is present in the Lp1 IGS, but it seems that different sequences are also able to act as ARSs (e.g. Peng *et al.*, 1993). Interestingly, there is a sequence in the Lp1 IGS that matches 19 of the 21 bp consensus sequence (WAWTTDDWWWDHWGWHMAWTT) that Dobbs *et al.* (1994) identified in all the eukaryote replication origins they looked at (Figure 3.8A). However they considered a 20/21 bp match to be significant, and I have not analysed the IGS for the other features they proposed are also important for replication origins.

The ERC amplification and reintegration model can explain the rapid clustering of IGS length variants that was observed. It may also explain the rapid spread of the deletion and the restriction of IGS length variants. However it is important to note that it is not mutually exclusive with unequal crossing over or gene conversion, and unequal crossing over is the best explanation for the generation of the IGS length heterogeneity. Therefore I propose that all of these mechanisms are occurring simultaneously in the rDNA, leading to a complex interplay of factors that affect homogenisation. Dvorák (1989; Dvorák *et al.*, 1987) has also proposed a similar interplay of saltatory replication via ERCs with other more conventional homogenisation mechanisms.

Walsh (1987) argued that unequal crossing over was an unsatisfactory mechanism to account for homogenisation, as it would involve intrachromatid recombination leading to deletion of repeat units, and a process that includes deletion results in the repeat number tending to zero (gambler's ruin). There is some experimental evidence for a tendency to lose rDNA units (Butler and Metzenberg, 1989; Butler and Metzenberg, 1990; Dvorák, 1989) in recombination events. On these grounds an amplifying process might be predicted, and the very action of ERC amplification and reintegration is one of amplification of unit number.

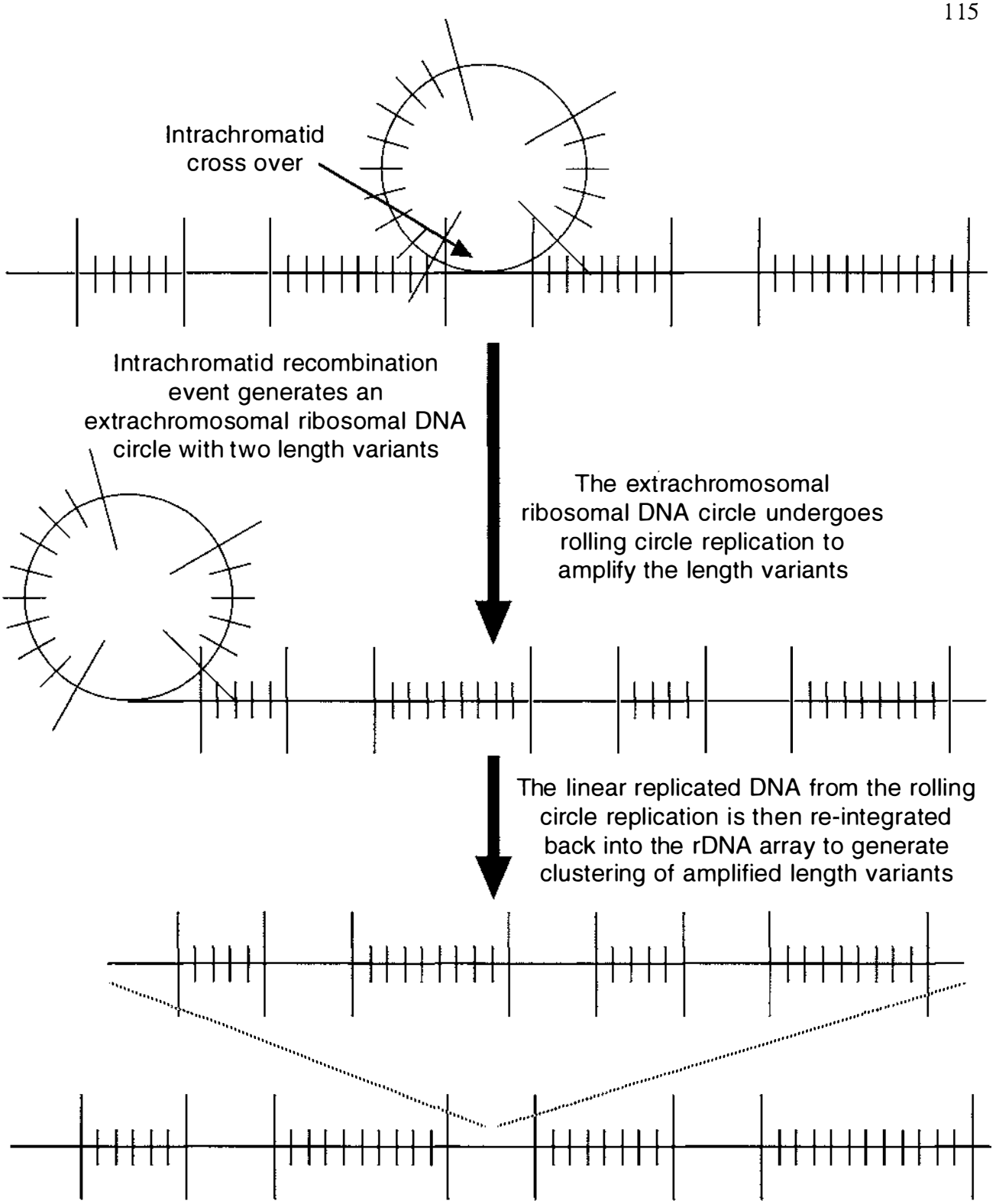



Figure 4.3 Extrachromosomal ribosomal DNA circle mechanism for generation of clusters of length variants

Diagrammatical representation of the rapid clustering of length variants in the rDNA by extrachromosomal ribosomal circle amplification and integration. The excision, replication and integration of the extrachromosomal circle is as described in the figure. The short horizontal lines the sub-repeats in the IGS, and the long horizontal lines delineate the IGS from the *rrn* genes. The linear replicated DNA molecule may circularise before reintegration into the rDNA. See text for further details. Diagram not drawn to scale.

The best-characterised ERCs are those that are amplified in the oocytes of insect and amphibian species, especially *Xenopus* (Hadjiolov, 1984; MacGregor, 1982). Here vast numbers are produced, presumably to cope with the high levels of protein synthesis that occur in the early embryo. It appears that the amplified rDNA units are largely representative of those in the chromosome, but the relative proportions can differ (Wellauer *et al.*, 1976b). The key point here is that extrachromosomal rDNA copies are not just a faithful representation of what is present on the chromosomes. They have also been found somatically, often alongside extrachromosomal circles of other repeated sequences (Hadjiolov, 1984; Pont *et al.*, 1987; Rochaix and Bird, 1975). Finally, Sinclair and Guarente (1997) showed that the aging phenotype in yeast was caused by the production of high numbers of ERCs in old yeast cells, which results in fragmentation of the nucleolus. They showed that it was not specifically the rDNA itself that gave the aging phenotype, but any extrachromosomal element that contained an equivalent ARS. It is a provocative idea that homogenisation in the rDNA via the generation of rDNA ERCs leads to the aging phenotype in yeast, but this is purely speculative.

ERCs are relatively common, and their integration back into the genome seems likely given the ease with which plasmids can be targeted to the rDNA, at least in the fungi. In fact, excision and integration of ERCs to and from the rDNA array in *S. cerevisiae* can be regulated by controlling the levels of topoisomerase I and II. Repressing expression of these proteins results in excision of ERCs and subsequent expression of either one results in reintegration of the ERCs back into the array (Kim and Wang, 1989). In this light it is interesting that topoisomerases have been found to suppress recombination within the rDNA (Christman *et al.*, 1988; Gangloff *et al.*, 1994).

### 4.5.3 THE NUCLEOLUS AS A REGULATORY CENTRE FOR rDNA HOMOGENISATION

The question of the role of the nucleolus in homogenisation has received almost no attention to date. The physical separation of the rDNA from the rest of the genome would naturally suggest that processes, such as homogenisation, that affect the rDNA could be isolated from the rest of the genome and could be under very tight control. It would make sense from a selectionist standpoint that the homogenisation of the rDNA would be a highly controlled, specialised process rather than left to the general turnover mechanisms at large in the genome. It has been suggested that interchromosomal homogenisation is much more likely if separate rDNA loci form a single nucleolus than if they each form independent nucleoli (Arnheim 1983 ; D.M. Hillis, personal communication). This would imply that the nucleolus has a central role in homogenisation. Indeed in wheat where the rDNA loci form independent nucleoli (Appels and Honeycutt, 1986), there appears to be some interlocus homogenisation occurring, but it is inefficient compared to homogenisation within the arrays (Dvorák, 1989).

Support for this idea also comes from work done on recombination and especially gene silencing in the rDNA. Genes for both these processes have been found that appear to act specifically, or

at least differentially, in the nucleolus (e.g. Bryk *et al.*, 1997; Cavalli *et al.*, 1996; Christiansen *et al.*, 1987; Christman *et al.*, 1988; Gottlieb and Esposito, 1989; Gottlieb *et al.*, 1989; Keil and McWilliams, 1993; Ozenberger and Roeder, 1991; Smith *et al.*, 1998). Although we still know little about how these processes are controlled and the precise role of the various gene products, a picture seems to be emerging that implicates chromatin structure as a central feature of the rDNA in controlling both transcription and recombination (Cavalli *et al.*, 1996; Dammann *et al.*, 1995; Gottlieb and Esposito, 1989; Kulkens *et al.*, 1992; Muscarella *et al.*, 1987; Thompson and Flavell, 1988). This fits in well with the little understood lack of transcription and magnification of rDNA units that contain insertion elements in *D. melanogaster* (Hadjiolov, 1984; Sharp *et al.*, 1983; Terracol, 1987). If these interrupted rDNA units were somehow recognised and maintained in a more repressive chromatin structure than the other repeats, this would repress their transcription and possibly their magnification as well.

Thus results from a number of fields are beginning to build a comprehensive picture of the dynamics of the rDNA in the context of the nucleolus. This picture is characterised by differential chromatin structure for different rDNA repeats within the nucleolus, and perhaps this controls the interactions of other regulatory proteins. These structures allow the cell to control the number of rDNA units, their transcriptional activity, their recombination and I propose their homogenisation as well. I have proposed that extrachromosomal amplification of rDNA circles plays a role in this homogenisation, with it too being under the control of the nucleolar architecture. The nucleolar architecture protects, but does not fully repress, the rDNA from general genomic turnover mechanisms, such as unequal crossing over and gene conversion. Indeed they are likely to play an important role in spread of variants in a cohesive manner through sexually reproducing populations (Ohta and Dover, 1984), in facilitating interlocus homogenisation and perhaps in dynamically altering transcriptionally-important repetitive variants for selection to act upon (Cluster *et al.*, 1987; Petes and Hill, 1988; Reeder *et al.*, 1983). The unequal crossing over/gene conversion paradigm of homogenisation is unable to explain all the results presented in this study. Given the complex and often contradictory results that have dogged researchers attempting to understand the molecular basis for homogenisation, I believe that a saltatory replication model of homogenisation by ERCs merits consideration alongside the unequal crossing over/gene conversion model, and that experimental testing of the model may quickly establish its validity or otherwise.

## 4.6 THE IGS LENGTH HETEROGENEITY ARISES THROUGH HYBRIDISATION

The extraordinary length heterogeneity in the Lp1 IGS seems to be a consequence of the hybridisation event, as neither of the two species representing the progenitors of Lp1 show such length heterogeneity. However it does not seem to be an outcome of hybridisation *per se*, as other hybrids from different hybridisation events that we have looked at do not display such IGS

length heterogeneity (Figure 3.11). Rather, there appears to have been a loss of control in the maintenance of length homogeneity as a result of the hybridisation in Lp1. The rate of recombination in multigene families such as the rDNA differs from that found in the genome in general (Christman *et al.*, 1988; Gottlieb and Esposito, 1989; Gottlieb *et al.*, 1989; Keil and Roeder, 1984; Ozenberger and Roeder, 1991) and mutations in some genes give recombination phenotypes that differ for repeated and non-repeated DNA (reviewed in Klein, 1995; Petes and Hill, 1988). Furthermore I have proposed that copy number control in the rDNA is under tight regulatory control (Section 4.5.1). Thus loss of homogeneity control could be the result of a disruption in an rDNA-specific recombination process that does not result in genome-wide perturbations.

The nature of this disruption of control is not known, and our lack of biochemical and genetic knowledge about the mechanisms of homogenisation only allows general speculations. These fall into three categories: (1) loss of alignment control of the 111/119 bp sub-repeats, allowing misalignment which would generate the length heterogeneity. This has been modelled by Williams and Strobeck (1985) who show that small changes in the proportion of unequal versus equal cross overs can have dramatic effects on the number of length variants maintained in an array; (2) loss of control of a maximum length for the 111/119 bp sub-repeat array. Stephan and Cho (1994) established a theoretical base for a maximum size limit for a repeat array which is required for its maintenance. Similarly Williams and Strobeck (1985) and Hood *et al.* (1980) used a maximum size limit in their calculations. Lyckegaard and Clark (1991) modelled regulation of rDNA unit number using a variety of parameters and found a number of ways the parameters could be manipulated to fit the experimental data. The IGS from several species contains enhancer elements and/or promoter duplications, and several studies have concluded that there is a window in the number of such sub-repeats in an IGS that gives a high rate of transcription, above and below which rDNA transcription efficiency decreases (Coen and Dover, 1983a; Lohe and Roberts, 1990; Reeder *et al.*, 1983). This is the best experimental evidence for a limit on sub-repeat array size. The control of rDNA unit copy number in *N. crassa* (Rodland and Russell, 1982; Russell and Rodland, 1986) and *S. cerevisiae* (Kaback and Halvorson, 1977) is also evidence for a copy number detection mechanism, but the molecular details for such a mechanism are entirely lacking; and (3) alteration in the relative balance of gene conversion and unequal crossing over in favour of crossing over, destabilising control of length.

All of these various types of disruptions could occur through loss of a gene via chromosome expulsion during hybridisation, gene doubling, or novel interactions between the two genomes analogous in a sense to the phenomenon of hybrid dominance (Reeder and Roan, 1984). If the latter is the case, then the generation of the length heterogeneity may have been a specific response to the genome duplication (Flavell, 1985). In this scenario, the process that resulted in the length heterogeneity is no longer active, but once generated the length heterogeneity is a natural target for unequal crossing over to perpetuate and perturbate as part of the "housekeeping" homogenisation process. In this sense it is interesting that all of the *Neotyphodium* hybrids to

date contain only a single type of rDNA sequence (Schardl *et al.*, 1991), and that cohabitation of dissimilar rDNA sequences types in a genome for long periods of time is unusual. As discussed in detail in Section 4.11 it seems that, for whatever reason, the genome usually does not tolerate substantially different rDNA types, and resolution of this situation (e.g. chromosome loss, interlocus homogenisation) may result in unusual "side effects". Drastic perturbations in methylation and non-coding repetitive elements upon hybridisation in the wallaby have been documented (Waugh O'Neill *et al.*, 1998).

## 4.7 A GROUP I INTRON IN THE 28S *rrn* GENE

Section 3.2 documented the discovery of a group I intron in the 3' portion of the 28S *rrn* gene of Lp1 and its progenitors, Lp5 and E8. A surprising feature of this group I intron is that it contains the *Sal*I restriction site that has been exploited in this study to document the extraordinary length heterogeneity seen in Lp1. The fact that only a single hybridising band is observed when a *Sal*I genomic digest is hybridised with a probe consisting of the coding region of the rDNA (Figures 3.3 and 3.4) suggests that this group I intron is present in every rDNA unit in these fungi (within the limits of detection). Furthermore, when *Sal*I digests of various other putative *Epichloë* endophyte hybrids were probed with the IGS probe, the hybridising bands were mainly less than the size expected for a whole rDNA unit (Figure 3.11), which suggests that many, if not all, of the *Epichloë* endophytes contain this group I intron. Also consistent with this, group I introns have been found to be widespread in both the 18S and 28S *rrn* genes of various clavicipitaceous fungi that are pathogens of insects, and these introns seem to be vertically inherited (N. Nikoh and T. Fukatsu, personal communication).

### 4.7.1 SEQUENCE COMPARISON OF THE LP1 GROUP I INTRON

Figure 3.15 shows an alignment of the Lp1 group I intron with group I introns that are most similar as determined by a BLAST search (Altschul *et al.*, 1997), and the other most similar group I intron inserting into the 28S *rrn* gene. Also shown in the alignment is the group I intron from *Beauveria brongniartii*, not available in the database, which is very similar to the Lp1 group I intron (C. Neuvéglise, personal communication).

Several investigators have proposed that group I introns that insert into the same site are phylogenetically related - they share common ancestry (Johansen *et al.*, 1992; Tan, 1997; Van Oppen *et al.*, 1993). It was surprising therefore that only two of the most similar introns to that in Lp1 inserted in the 28S *rrn* gene, with the other three all inserting into the 18S *rrn* gene. The *Gaeumannomyces graminis* intron T and the *B. brongniartii* intron both insert into the same site in the 28S *rrn* gene as that for Lp1, as would be expected. However the only other group I intron available in the database that also inserts into this site (*Naegleria* sp. intron 2) is not at all related in primary sequence (45% identity), a feature which the investigators who characterised

this intron also noted (De Jonckheere and Brown, 1998). It is clear that not all group I introns that insert into the same site are phylogenetically related and similarly some introns that insert into different sites are closely related (see also Hibbett, 1996). Sequence conservation without conservation of insertion position is evidence that group I introns are able to move into new sites within a gene and into new genes, a phenomenon reminiscent of transposition (Hibbett, 1996; Liu *et al.*, 1992; Nishida and Sugiyama, 1995; Sjamsuridzal and Sugiyama, 1998; Sogin *et al.*, 1986; Turmel *et al.*, 1993).

The pairing elements that I have predicted for the Lp1 group I intron (Figure 3.15) are very similar to those inferred for all other group I introns, with the P, Q, R and S elements pairing to form the catalytic core. This is further demonstration that this 28S *rrn* insertion element is indeed a group I intron, and suggests that the secondary structure fits the classical group I intron secondary structure (Cech, 1990). The putative internal guide sequence shown in Figure 3.16 does not fit with the classical internal guide sequence structure (Davies *et al.*, 1982) in that the P1 and P10 pairing elements at the 5' end of the intron do not seem to be spaced far enough apart to allow simultaneous pairing. The role of the P10 pairing as part of the internal guide sequence in the intron-splicing reaction is somewhat controversial (Cech, 1988; Woodson and Cech, 1989). If it does play a role, the P1 and P10 may pair sequentially through the splicing reaction, which would obviate simultaneous pairing. However there is evidence that the P1 and P10 elements compete for pairing, which implies simultaneous pairing (Thompson and Herrin, 1994).

### 4.7.2 MOBILITY OF GROUP I INTRONS

Several group I introns contain an open reading frame (ORF) that encodes an endonuclease. This endonuclease cleaves DNA in a sequence-specific manner, with the recognition site corresponding to the insertion site of its host intron (Lambowitz and Belfort, 1993). Cleavage of the DNA by this endonuclease results in a DSB break which is repaired by a gene conversion event, and this copies the intron from an intron-containing rDNA unit to the cleaved, intron-less rDNA unit, a phenomenon known as intron homing.

One form of mobility displayed by group I introns is horizontal transfer, the indendent transfer of introns between species. Group I introns that insert into nuclear *rrn* genes do so in highly conserved regions. In the last few years most studies that addressed the question have concluded that group I introns have been transferred horizontally multiple times (e.g. Cho *et al.*, 1998; Hibbett, 1996; Johansen *et al.*, 1992), although it is important to note that they are also able to transmitted vertically for long periods of time (e.g. Bhattacharya *et al.*, 1994). Horizontal transfer is obviously facilitated by the presence of these introns in conserved regions of the *rrn* genes, especially for those that encode an endonuclease. The patchy distribution of the ORFs that encode these endonucleases along with their apparent ability to move in and out of their host intron (Mota and Collins, 1988) suggests they are labile elements of group I introns. Therefore many of these horizontal transfer events may have been facilitated by endonuclease activity, with

the ORFs subsequently being lost to explain the lack of ORFs in many group I introns. It is not known what vectors (or other factors) are involved in the horizontal transfer of these introns.

As noted above, the presence of dissimilar introns in the same insertion site and especially the presence of related introns in dissimilar insertion sites is evidence for transposition, the other form of mobility displayed by group I introns. Rather than involve transferring the intron to the same site in a gene which lacks the intron like homing, transposition involves transferring the intron to a new site, perhaps even to a new gene. The question becomes what is the mechanism for transposition, as the endonucleases with their long recognition sites that are specific for the site of intron insertion are unlikely to be involved?

The finding that some of the most similar group I introns to that in Lp1 are inserted in the 18S *rrn* gene is evidence for transposition of this intron having occurred. The data I have presented provide no direct evidence for (or against) horizontal transfer of this intron, although the finding of a very similar group I intron in other clavicipitaceous fungi argue for long-term vertical transmission (N. Nikoh and T. Fukatsu, personal communication). Several authors have proposed an RNA-mediated mechanism for group I intron transposition (Belfort, 1993; Lambowitz, 1989; Turmel *et al.*, 1993). The generalised scheme of events is that an intron that has been spliced out of an RNA transcript splices back into another transcript at a different site or into a different transcript, in a reaction that is a reversal of the splicing action. There is experimental support for such reverse splicing (Mohr and Lambowitz, 1991; Thompson and Herrin, 1994; Woodson and Cech, 1989). The intron-containing transcript is then reverse transcribed, and integrates back into the genome. Evidence for reverse transcription comes from so-called processed pseudogenes that have characteristics of cDNA (Baltimore, 1985; Jagadeeswaran *et al.*, 1981), the reverse transcriptases present in retrotransposons (Derr and Strathern, 1993; Lathe and Eickbush, 1997) and the reverse transcriptases that some group II introns encode (Michel and Lang, 1985). That reverse splicing and reverse transcription are both likely to be rare events means that transposition is very likely to occur into RNA transcripts that are highly abundant, and indeed genes that host group I introns do produce abundant RNA transcripts (see Thompson and Herrin, 1994; Turmel *et al.*, 1993). Also of interest, Turmel *et al.* (1993) found that the majority of group I intron sites in *rrn* genes were in regions that are predicted from three dimensional models to be on the exterior of the ribosome, also supporting an RNA-mediated transposition mechanism. However *in vitro* work has suggested that reverse splicing is inhibited by RNA secondary structure (Mohr and Lambowitz, 1991; Thompson and Herrin, 1994; Woodson and Cech, 1989), and it is still not obvious why phylogenetically-conserved sites are favoured. Clearly there is much to understand about the mechanism of group I intron transposition.

The reverse splicing reaction requires at least the 5' end of the intron to pair with the 5' half of the insertion site (Woodson and Cech, 1989), with the requirement for 3' pairing less-established (Thompson and Herrin, 1994). I have checked the insertion sites of the Lp1 and 18S *rrn* group

I introns in the secondary structure of the 18S and 28S *rrn* genes, but these sites do not resemble each other more than other sites in these genes. It is possible that there is some sort of similar feature in the tertiary *rrn* structures that is not obvious from the secondary structure. It would also be possible for the internal guide sequence from one intron to fortuitously pair with other DNA sequences, thus allowing reverse splicing, especially given that G-U pairing can occur. I have not checked the putative internal guide sequences from the 18S *rrn* introns for the ability to cross-pair with the Lp1 intron insertion site. There is very little similarity in this region of these introns now (Figure 3.15), indicating that sequence changes will have obliterated any pairing similarities that may have existed at the time of transposition. On a cautionary note, while the 18S *rrn* introns are more similar over the whole intron to the Lp1 intron, the *C. albicans* intron in the 28S *rrn* gene is equally as similar to the Lp1 intron when just the P, Q, R and S regions are compared. It seems likely that the other regions are also phylogenetically important, but it is possible that the similarities outside the P, Q, R and S regions are due to high levels of homoplasy.

### 4.7.3 SPREAD OF THE GROUP I INTRON BY HOMOGENISATION

The finding of a similar group I intron in other clavicipitaceous fungi (N. Nikoh and T. Fukatsu, personal communication) suggests that if this group I intron has been horizontally transferred into the *Epichloë* endophytes, it may have done so very early in the evolution of the clavicipitaceous endophytes. However sequences of these introns are not available to test this. The group I intron in Lp1 does not encode an ORF, and therefore is presumed to be immobile. A single transfer event therefore implies that homogenisation has spread this intron throughout the rDNA array, or if an ORF was initially present and since has been lost, homogenisation has spread the ORF-less form of the intron throughout the rDNA. The fact that there are multiple rDNA loci in these endophytes (Section 3.3.2) suggests that interlocus homogenisation has also occurred. However, the origin of the multiple rDNA loci in these endophytes is unknown, and it is possible that at the time of the horizontal transfer of the intron there was only a single locus, which has subsequently been amplified.

## 4.8 *EPICHLOË* ENDOPHYTES CONTAIN MULTIPLE rDNA ARRAYS

Section 3.3 clearly demonstrates that the hybrid Lp1 and *N. lolii* LpTG-1 (Lp5) both contain at least five rDNA arrays. This was highly unexpected, as in the fungi for which the number of rDNA clusters has been reported, a maximum of two arrays has been found (Viaud *et al.*, 1996). The technical difficulties associated with determining rDNA locus number means that the situation is not known for most fungi, and it remains to be seen how general the presence of multiple rDNA loci in fungi is.

The origin of the multiple rDNA loci in these endophytes is unknown. I have not tested a wide range of *Epichloë* endophytes to determine how universal this feature is within this complex. This may prove difficult, as in this study I have been unable to adequately resolve the chromosomal bands from the sexual *E. typhina* E8 isolate. The reasons for the multiple rDNA loci are not clear. The propensity of these endophytes to hybridise (in evolutionary time) suggests that hybridisations very early in the evolution of the group may have generated the multiple loci, but once again we are hampered by the lack of resolution of E8 in assessing this possibility. Alternatively chromosomal translocations, with breakpoints within the rDNA array, seeding new arrays may be the source of the multiple loci (Adam *et al.*, 1991; Dubcovsky and Dvorák, 1995). Again the finding that the rDNA array is a fragile site in *N. crassa* (Butler, 1992) may be relevant.

### 4.8.1 THERE ARE BOTH MAJOR AND MINOR rDNA LOCI IN THE *EPICHLOË* ENDOPHYTES

One feature of the CHEF Southern blotting is the different hybridisation signal strengths shown by the various chromosomal bands that hybridise to the ribosomal probe (Figure 3.18B). Similarly, attempts to resolve the separate rDNA arrays resulted in two types of hybridisation - hybridisation to discrete, high molecular weight bands, and hybridisation to lower molecular weight smears (Figure 3.21B). Together these results indicate that these endophytes contain both major and minor rDNA loci (Dubcovsky and Dvorák, 1995) representing relatively many and few rDNA units in the array respectively.

The minor loci do not resolve as discrete bands on the gel, despite the *S. cerevisiae* chromosomes in the same part of the gel resolving well. Perhaps this smear of hybridisation is indicative of a large amount of length heterogeneity, even though there are likely to be less rDNA units in these minor loci. This length heterogeneity could be length heterogeneity of the IGS regions, but this seems unlikely as the smear of hybridisation is seen with both Lp5 and E8 which do not show the IGS length heterogeneity of Lp1. This study gives no indication whether the IGS length heterogeneity observed in Section 3.1.2 is restricted to one or a few of the rDNA loci, or is occurring in all of them. It is not at all clear why there should be heterogeneity in the minor loci, whilst the major loci are of a relatively stable size, and there remains the possibility that the smearing is an artefact of the pulsed field separation. It is also not known why in some organisms there are distinct major and minor loci.

## 4.9 CLONING AN rDNA JUNCTION

Significant delays were experienced during the attempt to clone rDNA junction fragments, primarily due to PCR failure. Once these delays were resolved, I obtained several potential junction fragments. Initial PCR screening of these potential fragments suggested that none were

true junction fragments, however subsequent sequencing of one of the clones revealed that it could, in fact, be a junction fragment. The sequencing did not resolve why the PCR did not work - this is likely to have been the result of further PCR problems. Sequencing of both ends of this single clone showed that one end was homologous to the 3' end of the 28S *rrn* gene, and the other end showed no homology to any known sequence. This is the pattern expected for a junction fragment, however the clone was from Lp5, and the sequence of much of the IGS in Lp5 differs from that from Lp1, and is not known. Therefore it is not possible to rule out this fragment being composed entirely of IGS (and 28S *rrn*) sequence, but I would expect some IGS homology to remain between Lp5 and Lp1. If this is indeed a junction fragment, then the junction point occurs in the IGS. This is consistent with other ribosomal junctions that have been characterised, the majority of which also occur in the IGS (Copenhaver and Pikaard, 1996a; Worton *et al.*, 1988; Zamb and Petes, 1982 but see Adam *et al.*, 1991; Ellis *et al.*, 1986). The delays in obtaining this junction fragment meant I did not have time to characterise the clone to confirm that it is a true junction fragment, and to further analyse it.

The fact that this clone may be a junction fragment makes it highly probable that some of the other fragments I obtained are also junction fragments. Given that interlocus homogenisation of the rDNA appears to have occurred extensively in Lp1 (see Section 4.11), these junction fragments would be extremely interesting to analyse. Key points are whether the junction fragments retain their parental origin or are converted along with the rDNA; whether the junction breakpoint in the rDNA is conserved between different loci and between the two progenitors; whether there is evidence of gene conversion tracts extending a short way into the junction regions; and whether there is some scrambling of repeat sequences at the junctions or other evidence for different mechanisms of homogenisation.

## 4.10 CHARACTERISATION OF THE 5S rRNA GENES

The results presented in Section 3.4.2 demonstrate that the 5S rRNA genes are organised as dispersed copies rather than in tandem clusters as is most commonly found in other eukaryotes (Hadjiolov, 1984). To date reports of a dispersed arrangement of the 5S rRNA have only been made from *N. crassa* (Selker *et al.*, 1981); *Ascobolus immersa* (Goyon *et al.*, 1996), *Cochliobolus heterostrophus* (Garber *et al.*, 1988); *Schizosaccharomyces pombe* (Mao *et al.*, 1982; Tabata, 1981); and *Acanthamoeba castellani* (Zwick *et al.*, 1991) amongst the eukaryotes that have tandem arrays of their rDNA. There have been reports of organisms having some dispersed 5S rRNA copies (orphons; Childs *et al.*, 1981) in their genome alongside 5S rRNA tandem arrays (Belkhiri *et al.*, 1992; Reddy *et al.*, 1986 see also Drouin and Moniz de Sá 1995).

### 4.10.1 STRUCTURE OF THE 5S rRNA GENE

Extrapolating from the primer binding sites determined from the 5S rRNA genes of other organisms, the sequence obtained from these *Epichloë* endophytes indicates that the 5S rRNA genes are 119 bp in length, as expected for a typical 5S rRNA gene. The sequences of the 5S rRNA genes are virtually identical between the three *Epichloë* endophytes, and the implications of this are discussed in depth in Section 4.10.3. Figure 3.26B shows an alignment of the 5S genes from these endophytes as well as those from *N. crassa, A. castellani, X. laevis* and *Rattus norvegicus.*

The internal control region, which is the primary regulator of 5S rRNA transcription initiation (Bogenhagen *et al.*, 1980; Sakonju *et al.*, 1980 although see Morton and Sprague 1984; Selker *et al.*, 1986) is thought to consist of the A box, C box and intermediate elements (Pieler *et al.*, 1987), and these are shown on the alignment in Figure 3.26B. The A and C boxes are believed to be the regions to which transcription factors TFIIIC and TFIIIA respectively bind (Geiduschek and Tocchini-Valentini, 1988). As can be seen, these A and C boxes show relatively high levels of conservation, with the intermediate element between them showing less homology. There are also regions in the rest of the 5S rRNA gene that are conserved as well, and these presumably relate to areas that are important for ribosomal function. The sequence of the 5S rRNA genes from Lp1, Lp5 and E8 are entirely consistent with the generalised features of 5S rRNA genes from other eukaryotes.

### 4.10.2 THE 5S rRNA GENES AND THE rDNA LOCI ARE FOUND ON THE SAME CHROMOSOMAL BANDS

The fact that the 5S rRNA genes are located on the same chromosomal bands as the rDNA loci was unexpected, and is in contrast to the situation in *N. crassa* where the 5S rRNA genes are present on at least six of the seven chromosomes (Metzenberg *et al.*, 1985). This is not novel, as the same situation has been found in various cereals. The 5S rRNA genes are located on two of the three chromosomes that the rDNA loci in wheat are located on, and on the rDNA-containing chromosome in rye (Appels *et al.*, 1980), despite the fact that the two gene families are not linked together. However in these cereals the 5S rRNA genes are organised as tandem clusters, and one would predict that some dispersed copies are able to spread to other chromosomes with a greater probability than an entire multigene locus is, or put another way, restriction of loci to a small number of chromosomes intuitively seems more likely than restriction of many dispersed (and presumably independent) genes. The presence of both types of ribosomal gene on the same chromosomes is suggestive of a means of regulatory control, as both types of product are required in equimolar amounts for ribosomal biogenesis (Gerbi, 1985). Why only these two highly divergent species (of those with unlinked 5S rRNA and rDNA genes) have this organisation is not known. The chromosomal location of the 5S rRNA and rDNA genes has not been studied in many species, and once more have been analysed a coherent pattern may emerge.

The discovery of 5S rRNA genes linked to the rDNA in some organisms initially led to the suggestion that this was the ancestral state, and the linkage has since been lost in most lineages (Belkhiri *et al.*, 1992; Gerbi, 1985). However Drouin and Moniz de Sá (1995) have presented convincing evidence for the opposite - that the 5S rRNA genes have become linked to the rDNA, and indeed have repeatedly become linked to other tandemly-repeated gene families as well, in a number of independent events. Unfortunately the place of dispersed 5S rRNA genes in the evolutionary dynamics of the 5S multigene family has not been considered important. The results reported here, I believe, point to dispersed genes being important in the transitions of 5S rRNA gene organisations. States of organisation identified to date are: dispersed and not associated with the rDNA chromosomes; dispersed and associated with the rDNA chromosomes; tandemly arrayed and not associated with the rDNA chromosomes; tandemly arrayed and associated with the rDNA chromosomes; and linked to the rDNA array (or another multigene family). It seems that 5S rRNA genes can move relatively freely between the different states, and this may relate to the genome dynamics of the system in question. Interestingly, one of the mechanisms suggested by Drouin and Moniz de Sá (1995) that might be responsible for this 5S rRNA mobility is the excision and reintegration of 5S rRNA extrachromosomal circles. Mobility of rDNA arrays has also been suggested by Adam *et al.* (1991) and Andronico *et al.* (1985).

### 4.10.3 CONCERTED EVOLUTION OF THE 5S rRNA GENES

The concerted evolution of the 5S rRNA genes in organisms, such as these *Epichloë* endophytes, that have dispersed copies presents a problem. The conventional unequal crossing over model is not tenable amongst dispersed repeats, as it would result in major chromosomal rearrangements which are likely to be lethal. It is also questionable whether these genes would be able to pair for an exchange, given their small size. The saltatory replication model for homogenisation which I presented in Section 4.5 is obviously not possible in this system either.

Comparatively little work has been done on the concerted evolution of dispersed repeats. Probably part of this can be assigned to the fact that homogenisation is not particularly efficient compared to that operating in tandemly arrays, as evidenced by a sub-family structure seen in such repetitive families (Brown and Dover, 1981; Selker *et al.*, 1981). Transposition and gene conversion are the two mechanisms invoked to explain homogenisation of dispersed repeats. The most detailed study of dispersed 5S rRNA gene evolution was that performed by Morzycka-Wroblewska *et al.* (1985) on the genes in *N. crassa*. They concluded that transposition was unlikely to be responsible for the homogenisation of the genes, as the genomic locations of the 5S rRNA genes were relatively stable through evolutionary time. Therefore they proposed gene conversion as a likely mechanism for homogenisation of these dispersed repeats. Interestingly whilst the genes remained highly similar, their flanking sequences were able to diverge. This suggests that only the coding regions are homogenised. In this light Morzycka-Wroblewska *et al.* (1985) suggest that the gene conversion event could be RNA-mediated, and Drouin and Moniz de Sá (1995) consider this as a mechanism for the apparent mobility of 5S rRNA genes in

the genome. Contrary to this Morzycka-Wroblewska *et al.* (1985)also found an upstream region of a few base-pairs (which they suggested was involved in transcriptional regulation) which was conserved. The level of divergence around this area was not sufficient to distinguish homogenisation from drift, however.

One obvious question is how does pairing occur between such small repeats to allow gene conversion? Sugawara and Haber (1992) found that gene conversion only requires around 80 bp of homology in *S. cerevisiae*, although this was for repeats on the same chromosome located close to each other. Jinks-Robertson *et al.* (1993) found that recombination between repeats located on separate chromosomes required around 250 bp of homology, again in *S. cerevisiae.* However the 5S genes are only 160 bp including the upstream region. Munz *et al.* (1982) and Amstutz *et al.* (1985) showed that gene conversion events occurred between widely dispersed tRNA genes in *S. pombe*, and proposed that this was responsible for the concerted evolution of these genes. The events occurred between dual tRNA genes, so there was 200 bp homology between the sequences undergoing gene conversion. Gene conversion appeared to occur at a much lower frequency between single tRNA genes that were only 100 bp in length. Overall, these results suggest that gene conversion could be an adequate mechanism for maintaining homology between dispersed genes such as the 5S rRNA genes, especially given that homogenisation does not appear to occur as efficiently in dispersed copies as it does in multigene arrays (Brown and Dover, 1981).

In this study the hybridising bands shown in Figure 3.27 correspond to the multiple 5S rRNA gene loci. The hybridising bands in the hybrid Lp1 are largely a composite of the bands from the two progenitors, suggesting that a majority of the 5S rRNA genes from each progenitor are present in the hybrid. Also Lp5 and its presumed sexual ancestor, *E. festucae*, share a number of hybridising bands. This indicates that the genomic regions into which the 5S rRNA genes are located are relatively stable. These results are in concordance with those of Morzycka-Wroblewska *et al.* (1985), suggesting that transposition is not responsible for the homogenisation of this gene family. However it is possible that the time since speciation of Lp5 and the hybridisation that formed Lp1 has not been sufficient for many transposition events to have occurred. Pavelitz *et al.* (1995) found a similar result for the tandemly-repeated snRNA U2 multigene family.

Comparison of the 5S rRNA gene sequences from the three isolates shows that they are extremely similar. These are the effective 'consensus' sequences of the 5S rRNA genes in each isolate, but it is important to note that they are the consensus sequences only for those products which are amplified by the primers used. If there is a sub-family structure in the 5S rRNA genes which includes differences in the priming site (s), then the other sub-families would not be amplified. The presence of ambiguous positions in the sequence suggests that these primers are amplifying polymorphic variants, and the fact that two of the three ambiguous positions are shared between all three isolates indicates that these polymorphic variants have been maintained

throughout the evolutionary divergence of these isolates. This indicates that the rate of homogenisation has not been great enough to allow fixation of one variant in most of the 5S rRNA copies of any of these endophytes, and this polymorphism must date back at least to the common ancestor of *E. festucae* and *E. typhina.* However accurate assessment of the rate of homogenisation of these dispersed 5S rRNA genes would require detailed characterisation of the level of heterogeneity in the 5S rRNA gene family. It would be interesting to clone a number of 5S rRNA genes and their flanking regions and compare these, both within an isolate and between the isolates. This may give information on the mechanism of homogenisation (e.g. gene conversion), and would reveal whether there was a sub-family structure to the 5S rRNA multigene family in these *Epichloë* endophyte isolates. Maintenance of polymorphic variants through evolutionary time and the presence of sub-family structure are both potential indicators that the rate of homogenisation is lower than that of a tandemly-repeated multigene family, such as the rDNA. If, as I proposed in Section 4.5.3, the nucleolus is an important regulator in the homogenisation of the rDNA, then the fact that the 5S rRNA genes are not transcribed in the nucleolus (Hadjiolov, 1984) may be another factor responsible for the lower homogenisation rate of the 5S rRNA genes.

## 4.11 INTERLOCUS HOMOGENISATION IN THE HYBRID, Lp1

Although individual methods for distinguishing between interlocus homogenisation and rDNA locus loss to explain only one type of rDNA sequence in the hybrid have not been conclusive in this study (Section 3.3), primarily because of the inability to adequately resolve the chromosomal bands from E8, the cumulative weight-of-evidence supports interlocus homogenisation. The large number of rDNA loci in at least Lp5 and Lp1 is surprising (Section 3.3.2). It also means that to rule out the interlocus homogenisation hypothesis it is necessary to invoke loss of all five rDNA loci originating from Lp5, but to maintain five rDNA loci from E8. Whether E8 has five rDNA loci is a moot point, but even if it did such a degree of biased loss of rDNA loci seems unlikely, especially considering the similarity in size of the rDNA-containing chromosomal bands in Lp5 and Lp1. It cannot be ruled out, however, as results from allopolyploids in some plants suggest that biased loss of rDNA loci can occur (Dvorák, 1989; Vaughan *et al.*, 1993). This loss may be associated with the transcriptional advantage that one rDNA type has over the other, a phenomenon known as nucleolar dominance (Reeder and Roan, 1984; Thompson and Flavell, 1988).

The most compelling evidence for interlocus homogenisation comes from the work done on the 5S rRNA genes. As discussed above, the 5S rRNA genes are located on the same chromosomal bands as the rDNA loci in both Lp1 and Lp5 (again the situation in E8 cannot be resolved). Thus to argue for rDNA loci loss, assuming that at least some of the chromosomal bands represent single chromosomes, all the rDNA-containing chromosomes from Lp5 must have been lost in the hybrid. Lp5 and Lp1 have many bands in common that hybridise to the 5S rRNA probe (Figure

3.27). Therefore these 5S rRNA genes have not been lost in the hybridisation. These 5S rRNA genes represent a type of marker for the chromosomal bands that contain the rDNA, and demonstrate that at least some of the rDNA-containing chromosomes are still present in Lp1. I propose that the presence of only one type of rDNA sequence in the hybrid, Lp1, is the result of interlocus homogenisation.

This demonstration of interlocus homogenisation of the rDNA in Lp1 raises a number of interesting points. The most obvious is that this homogenisation must have occurred extremely rapidly. Lp1 is believed to be a recent hybrid (Collett, 1994), and this is supported by the lack of differences between Lp1 and E8 in the regions of the IGS that I have sequenced (Sections 3.1.6 and 3.2.1). In the time since homogenisation it appears that several rDNA loci from Lp5 have become homogenised to E8 sequence. This homogenisation is extensive, as I have been unable to find any evidence for residual Lp5 rDNA units, within the resolution of the Southern blotting. Furthermore the homogenisation seems highly biased. Although I have been unable to determine the number of rDNA loci in E8, it is highly unlikely that there are more than five, and it would not be surprising given the hybridisation results if there were less than five.

The rapid rate of homogenisation and its biased nature may be connected. As mentioned above, in hybrids one rDNA type tends to have a transcriptional advantage over the other - nucleolar dominance, and there is a suggestion that this plays a role in the biased loss of rDNA loci. It may be that biased chromosome loss and biased interlocus homogenisation (Hillis *et al.*, 1991; Vaughan *et al.*, 1993; Wendel *et al.*, 1995) are two different ways of dealing with the same problem - ridding the cell of transcriptionally-inefficient rDNA loci. The transcriptional efficiency of one particular rDNA type may result in its rapid and biased homogenisation. A link between transcriptional activity and mitotic recombination is well documented, including recombination in the rDNA (Stewart and Roeder, 1989; Thomas and Rothstein, 1989; Voelkel-Meiman *et al.*, 1987). The basis for this link is not fully understood, but different chromatin structures are believed to be central (Bryk *et al.*, 1997; Gottlieb and Esposito, 1989; Wu and Lichten, 1994) and so this continues the theme I introduced in Section 4.5.3 on the role of chromatin structure in regulating the rDNA in the nucleolus. In the case of Lp1, the greater transcriptional activity of the E8-derived rDNA may have promoted a high level of recombinatorial activity (be it unequal crossing over, gene conversion, or ERC production) which resulted in replacement of the Lp5-derived rDNA, although it should be noted that there is no experimental evidence for E8 rDNA having a transcriptional advantage over Lp5 rDNA.

It is not known whether this interlocus homogenisation in the hybrid represents an accelerated rate in response to different rDNA types, or is the normal, background rate. If it is an induced form, the homogenisation may have also resulted in the appearance of significant length heterogeneity, as discussed in Section 4.6. On the other hand, there are a number of examples where interlocus homogenisation has not occurred or has only partially occurred (Waters and Schaal, 1996; see also Section 1.3.2). In many cases introgression is believed to be occurring,

which may tend to counteract the effects of homogenisation (Arnold *et al.*, 1988; Odorico and Miller, 1997; Rieseberg, 1991; Soltis and Soltis, 1991). Zhang and Sang (1999) propose that interlocus homogenisation, in plants at least, is driven by crossing over between loci, and therefore can only occur when the rDNA loci are located telomerically or sub-telomerically.

# CHAPTER FIVE

# CONCLUSIONS AND FUTURE DIRECTIONS

# 5.1 CONCLUSIONS

This study addressed mechanisms that are potentially responsible for the homogenisation of the ribosomal RNA genes, leading to the observed patterns of concerted evolution of these genes. This was performed in a fungal hybrid, for which the presumed progenitors have been characterised, thus allowing comparisons between these isolates. I looked at three aspects of the ribosomal DNA in this system:

(1) Length heterogeneity in the IGS region of the rDNA;
(2) Interlocus homogenisation of the rDNA in the hybrid;
(3) Evolution of the 5S rRNA genes in the hybrid.

The IGS region of the rDNA in the hybrid, Lp1, was found to vary widely in length within the genome, while such length heterogeneity was not present in the IGS of either progenitor. The length heterogeneity was shown to be a result of copy number variation of sub-repeats found within the IGS. This is evidence for unequal crossing over occurring in the rDNA, and suggests that unequal crossing over is a mechanism of homogenisation. However further characterisation of the IGS length heterogeneity in Lp1 produced results that are not predicted by an unequal crossing over model of homogenisation.

I have drawn two conclusions from these results: that the cross over site in the IGS is not random, but is specifically initiated at one side of the sub-repeat array that varies in length; and that other mechanisms are also involved in the homogenisation of the rDNA. Gene conversion is able to explain the restriction of the number of IGS length variants to a relatively small number and the lack of sub-repeat terminal exlusion which are not explained by an unequal crossing over model. However, neither gene conversion nor unequal crossing over is able to explain the rapid clustering of multiple copies of a single length variant, and I have proposed extrachromosomal ribosomal DNA circle amplification by rolling circle replication followed by reintegration into the rDNA array as a mechanism to generate this result, and so as a mechanism of homogenisation. Therefore the results I have presented suggest that homogenisation in the rDNA in Lp1 is achieved by a complex interplay of mechanisms, the molecular basis of which are all only poorly understood.

The hybrid, Lp1, has a genome that appears to be a composite of its two progenitors, Lp5 and E8. One exception to this is the rDNA - Lp1 appears to contain only E8-type rDNA. I tested the hypothesis that the presence of only one rDNA type in the hybrid is the result of interlocus homogenisation of the Lp5 rDNA arrays to E8 sequence, as has been found in lizard and cotton hybrids (Hillis *et al.*, 1991; Wendel *et al.*, 1995). Using pulsed field gel electrophoresis I demonstrated that both Lp1 and Lp5 contain at least five rDNA arrays, an unexpected result, as most organisms' rDNA is located on fewer arrays. I was unable to resolve the E8 chromosomal

bands, so a comparison to this progenitor was not able to be made. Resolution of whole rDNA arrays demonstrated that there are two types of array: high molecular weight arrays of a discrete size; and lower molecular weight arrays of a highly variable size. There appear to be two of the former type of array in the hybrid and both progenitors, whilst the number of the latter type of array cannot be estimated because of the extreme variation in size that these arrays exhibit.

I investigated the hypothesis that interlocus homogenisation of the 5S rRNA gene array in hybrids occurs at a much slower rate than that of the rDNA. However, I showed that the 5S rRNA genes in Lp1, and indeed most probably in all the *Epichloë* endophytes, are not arranged in a tandem array but are present as dispersed copies. This prevented me from testing the original hypothesis, as it is pointless to compare the homogenisation of a tandemly-repeated multigene family with a dispersed one. However I showed that the 5S rRNA gene profile of Lp1 was largely a composite of that from Lp5 and E8, and that the profile of Lp5 was very similar to that of its putative sexual progenitor, *E. festucae*. These results indicate that transposition is not likely to play a major role in the homogenisation of dispersed multigene families. Unequal crossing over and extrachromosomal circle amplification and reintegration are also unlikely to homogenise these dispersed repeats, thus gene conversion is the most likely mechanism for homogenisation of the 5S rRNA genes in these endophytes. Although I have been unable to test my original hypothesis that homogenisation of 5S rRNA gene arrays is slower than that in rDNA arrays, a slow rate of homogenisation of the 5S rRNA genes in these endophytes is suggested by the presence of variants that appear to be shared between Lp1, Lp5 and E8. Homogenisation may be slowed down in the dispersed 5S rRNA genes as a result of the small length of homology that is available for pairing, as a result of only a single mechanism (gene conversion) rather than an interplay of mechanisms bringing about homogenisation, as a result of dispersal of these genes in a relatively large amount of "genome space", or a combination of these. This slow rate of homogenisation does not necessarily apply to 5S rRNA gene homogenisation in general, therefore the original hypothesis remains untested.

Determination of the chromosomal location of these dispersed 5S rRNA genes by pulsed field gel electrophoresis showed they were present on the same chromosomal bands as the rDNA arrays. The significance of this finding is unclear at the moment, but it may relate to the mobility of the 5S rRNA genes and their apparent propensity to become associated with other multigene families (Drouin and Moniz de Sá, 1995). A fortuitous outcome of this experiment is that the 5S rRNA gene loci can be used as a form of chromosomal marker, demonstrating that at least some of the chromosomes from Lp5 that contain the 5S rRNA genes, and therefore rDNA loci, are present in Lp1. This result, in conjunction with the other results from the pulsed field gel electrophoresis work, builds a strong case for the lack of Lp5 rDNA sequence in Lp1 being a result of interlocus homogenisation. It is the first documented example of interlocus homogenisation occurring in fungi, and extends the generality of interlocus homogenisation found in a metazoan and a plant to the fungal kingdom. The generation of IGS length heterogeneity and the extensive interlocus homogenisation in the hybrid may be indicative of large-scale changes in the rDNA that are

induced by the presence of dissimilar rDNA arrays, perhaps as a result of their different transcriptional rates.

In the course of these studies I discovered a group I intron present in the 3' portion of the 28S *rrn* gene of Lp1. This must be present in Lp5 and E8, and is likely to be present in many of the *Epichloë* endophytes as it contains a *Sal*I site not normally present in the rDNA. This intron does not encode an endonuclease like several other group I introns, and may be an ancient feature of the clavicipitaceous fungi. There is evidence that this gene has been involved in transposition events to other *rrn* sites, possibly involving an RNA-mediated transposition mechanism.

The use of a hybrid organism in this study demonstrates the value of such systems for studying concerted evolution. The main advantages of the hybrid to come out of this study were the ability to show the original size of the rDNA IGS before the hybridisation event resulted in the extraordinary length heterogeneity seen in Lp1, and the ability to demonstrate interlocus homogenisation between the rDNA arrays from each progenitor in the hybrid. The latter result is unequivocal evidence for homogenisation being active in an asexual organism, and also implies that such homogenisation is rapid. The rDNA-DNA junctions can be studied in a hybrid where interlocus homogenisation has occurred to gain information on the homogenisation mechanism. I was not able to complete this line of investigation in this study, but comparison of such junction fragments before and after homogenisation (i.e. in the progenitors and the hybrid) is another substantial advantage of working with a hybrid system.

Recurring themes in this thesis on homogenisation in the rDNA are the regulation of rDNA unit copy number, and the role of transcription in various aspects of rDNA regulation. Although there is no direct evidence that regulation of rDNA unit copy number is mediated by the same mechanism (s) that bring about homogenisation, copy number variation is predicted by an unequal crossing over model of homogenisation, and control of rDNA copy number appears far from trivial. Transcription is known to influence recombination, and is implicated in rDNA unit and array loss. If so, then transcription is also a candidate for regulating homogenisation. Regulation of transcription itself seems to be mediated through chromatin structure, and the chromatin structure or at least the maintenance of it in the rDNA appears to be specialised. We therefore have a growing picture of the complex regulatory interplay of transcription, recombination, homogenisation, copy number control and perhaps other factors in the rDNA, mediated by chromatin structure and undoubtedly other regulatory proteins, that is organised around the nucleolus.

## 5.2 FUTURE DIRECTIONS

### 5.2.1 DETERMINATION OF HOMOGENISATION MECHANISMS

The most direct way currently available for determining the mechanism of interlocus homogenisation and therefore probably homogenisation in general is characterisation of the regions flanking the rDNA. Attempts to clone these junctions in this study were fraught with difficulty, but appear to have resulted in the cloning of one junction fragment. This suggests that cloning junction fragments with the methodology employed in this study is feasible, and that it should be possible to clone more junction fragments. Comparison of the junction fragments between Lp1 and its progenitors is likely to give information on the mechanisms that have brought about interlocus homogenisation in Lp1, and may also give definitive evidence for interlocus homogenisation. The system used in this study is an ideal one for this type of analysis, as the progenitors (or very close to them) of the hybrid, Lp1, are both known and therefore a direct comparison can be made between the state of the junction fragments before and after interlocus homogenisation.

### 5.2.2 TOWARDS AN ASSAY FOR HOMOGENISATION

An important step in attempting to "pin down" the molecular mechanism (s) of homogenisation and unravelling it from the complex interplay of events that occur in the nucleolus is the development of an assay for homogenisation. Having a system where carefully-controlled perturbations (such as genetic mutations) can be made and the effects on homogenisation can be monitored would allow direct tests of hypotheses concerning homogenisation to be performed. The only assay available for homogenisation has been observation of the behaviour of marker genes inserted into a repeat array (Gangloff *et al.*, 1996; Jinks-Robertson and Petes, 1993; Petes 1980; Szostak and Wu, 1980). However evidence suggests that a polymerase II-transcribed fragment of DNA inserted into the rDNA retains its own recombination pathway and dynamics, rather than behaving as part of the rDNA (Gottlieb *et al.*, 1989), reducing the utility of this approach. A key outcome of this thesis is evidence that the rate of homogenisation is much faster than most investigators have previously assumed. It is pointless attempting to assay homogenisation if it only occurs over evolutionary time, as we do not have a million years to wait for the results of an experiment. Homogenisation occurring over a small number of years is still not ideal, but it may be possible to artificially speed up the fixation time to experimentally-tractable periods. This would involve a system with preferably one rDNA array in which a variant is present in one or a very few copies, so the *Epichloë* endophytes are not ideal for this type of analysis. Assuming the variant is detectable, its spread can be monitored through time, and the time to homogenisation can be reduced by periodically selecting isolates that are closest to fixation. This is, in effect, attempting to remove part of the stochastic nature of homogenisation whilst not altering the underlying mechanism (s).

The homogenisation-like process that Muscarella and Vogt (1993) discovered may be an extremely rapid form of homogenisation, in which case it would be an ideal homogenisation assay. However I believe it is important to determine whether it is the result of the same mechanism that brings about "evolutionary" homogenisation, and this would probably require an assay of the type outlined above first. Questions as to whether the rate of homogenisation in one model system is applicable to other organisms can be answered by an in-depth assessment of the level of variation present in multigene families. The level of variation in a multigene family is the product of the rates of homogenisation versus mutation, or more accurately the rates of effective homogenisation versus mutation, assuming the number of genes in the family is constant. If the mutation rate is assumed to be constant (or is known), then the rate of homogenisation should correlate with the amount of variation present in the multigene family.

### 5.2.3 OTHER DIRECTIONS

Analysis of the 5S rRNA genes in these *Epichloë* endophytes showed that they are not tandemly arranged but are present as dispersed copies. Therefore my hypothesis that 5S rRNA gene arrays do not undergo interlocus homogenisation as rapidly as the rDNA arrays was not able to be tested, but I believe is still a valid hypothesis. Testing it would require finding a hybrid where interlocus homogenisation of the rDNA had occurred, and looking for evidence of 5S rRNA interlocus homogenisation. This could involve investigation of species which are suspected to be hybrids, yet only contain one form of rDNA.

The fact that the 5S rRNA genes are dispersed presents other interesting problems for the homogenisation of these genes. I have proposed that homogenisation of these dispersed repeats is less efficient than that of tandem repeats. To properly assess the validity of this hypothesis would require an assessment of the levels of variation present within the 5S rRNA genes versus the rDNA. Features such as a sub-family structure to the 5S rRNA genes and the persistence of polymorphisms through evolutionary time would also be indicative of slower homogenisation. It would also be interesting to determine whether there was evidence for homogenisation outside of the 5S rRNA genic region, or whether homogenisation acts strictly upon the gene.

This study demonstrated the presence of a group I intron in the 28S *rrn* gene of several *Epichloë* endophytes. Length variation between Lp5 and E8 in the *rrn*-containing part of the rDNA suggests that other group I introns may be present within these genes that are not so ubiquitously present in the *Epichloë* endophytes. It would be interesting to characterise any other group I introns in these endophytes to determine their relationship to the intron discovered in this study. They may help answer questions such as whether transposition of the intron has occurred in *Epichloë* endophyte lineage, or if there has been horizontal transfer of introns into the *Epichloë* endophytes. Finally analysis of the group I intron characterised in this study (and perhaps other introns) could be extended to other clavicipitaceous fungi to determine if long-term vertical transmission of this intron is a feature of this group of fungi.

# APPENDIX

# APPENDIX 1 PUBLICATION

Ganley, Austen R.D. and Scott, Barry 1998. Extraordinary ribosomal spacer length heterogeneity in a Neotyphodium endophyte hybrid: implications for concerted evolution. *Genetics* **150**, 1625-1637.

# APPENDIX 2 SEQUENCE OF THE POTENTIAL Lp5 rDNA JUNCTION FRAGMENT

Sequence of the 3.2 kb Lp5 potential nts1/nts11 junction fragment from Section 3.3.5.

A Sequence in from the rDNA IGS using the nts11 primer

```
  1  GGTTGGAGCC TTGCTTNAGT GATCTTTCAA GACCCTTGTA GAGCGCGGTA
 51  AGGTGTGGGC GGTTTTCTCG GGAGACTGCT TAAGGGACGT GCCAGACCCA
101  CNGGAAACCG TGCCGAATGC GAAGGACCTG CAGTCTGGAT CATTTGGAGG
151  GCGATTGAGT GCGGGAGGAA ATGCCCGTTA CGAGCCCGGT AGAATCACAC
201  AGTTAGGGGA TGAGTTGTCC TGTGTGGTTG TTCGAATAGG GAACATGAGC
251  TGGGTTTAGA CCGTCGTGAG ACAGGTTAGT TTTACCCTAC TGATGACCTN
301  GCCGCAATGG TAATTNAGCT TAATACGAGA GGAACCGCTG NTTCANATAA
351  TTGGTTTTTG CGGCTGTNCG ACCGGGCAGT GCCCGCNAAG CTACNATCTG
```

B Sequence in from the linker using the TAG primer

```
  1  CCTGAGGGGG AACCTCAAAC CGAGCAATCG GNTGGCTCAC AAGGTAGCAA
 51  AGAAGGCCTG CACAAAAGAC ACCTCCAGTG GCGCCGCAAC TCGCAGGAGA
101  GACACAATTT GTTGCACGAC CTGCCTGACG GCCTTGAGAT GCGAACCAGA
151  CCAGAATGGT ATATGGCAAA CCGATGAGAG ACCATGACGA GGCGAATCGA
201  AGAGGAATCA CGCGCCTACA TAGAGGATAA TAGCACACAC ATAGAAGACA
251  ACAGGGCACC CGCACGCATT GATGGCGCCC GCACGAGCAA CGACAGGGCA
301  CCCGCACGCA GNTACAGCGC CGGGACGGGC AACAACAGCG CACCCGCACG
351  CAGNTACAGC GCCGGGACGG ACAACAACAG CGCACCCGCA CGCATTGATA
401  GCGCCCGCCC ACACACCCAC AGCGCCGGCA CGGGCAACGA CAGCGCGCCC
```

# BIBLIOGRAPHY

Adam, R. D., T. E. Nash and T. E. Wellems, 1991. Telomeric location of *Giardia* rDNA genes. *Mol. Cell. Biol.* **11:** 3326-3330.

Aguilera, A., and H. L. Klein, 1988. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119:** 779-790.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

Amstutz, H., P. Munz, W.-D. Heyer, U. Leupold and J. Kohli, 1985. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell* **40:** 879-886.

Andronico, F., S. De Lucchini, F. Graziani, I. Nardi, R. Batistoni and G. Barsacchi-Pilone, 1985. Molecular organization of ribosomal RNA genes clustered at variable chromosomal sites in *Triturus vulgaris meridionalis* (Amphibia, Urodela). *J. Mol. Biol.* **186:** 219-229.

Appel, D. J., and T. R. Gordon, 1996. Relationships among pathogenic and non pathogenic isolates of *Fusarium oxysporum* based on the partial sequence of the intergenic spacer region of the ribosomal DNA. *Mol. Plant-Microbe Interact.* **9:** 125-138.

Appels, R., and J. Dvorák, 1982. The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. *Theor. Appl. Genet.* **63:** 337-348.

Appels, R., W. L. Gerlach, E. S. Dennis, H. Swift and W. J. Peacock, 1980. Molecular and chromosomal organization of DNA sequences coding for the ribosomal RNAs in cereals. *Chromosoma* **78:** 293-311.

Appels, R., and R. L. Honeycutt, 1986 rDNA: evolution over a billion years, pp. 81-131 in *DNA Systematics. Volume II: Plants*, edited by S. K. Dutton. CRC Press, Boca Raton, FL.

Arnheim, N., 1983 Concerted evolution of multigene families, pp. 38-61 in *Evolution of Genes and Proteins*, edited by M. Nei, and R.K. Koehn. Sinauer Associates, Sunderland, MA.

Arnheim, N., and M. Kuehn, 1979. The genetic behaviour of a cloned mouse ribosomal DNA segment mimics mouse ribosomal gene evolution. *J. Mol. Biol.* **134:** 743-765.

Arnheim, N., D. Treco, B. Taylor and E. M. Eicher, 1982. Distribution of ribosomal gene length variants among mouse chromosomes. *Proc. Natl. Acad. Sci. USA* **79:** 4677-4680.

Arnold, M. L., N. Contreras and D. D. Shaw, 1988. Biased gene conversion and asymmetrical introgression between subspecies. *Chromosoma* **96:** 368-371.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman and K. Struhl, 1987-1993 *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.

Baltimore, D., 1985. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* **40:** 481-482.

Barker, R. F., N. P. Harberd, M. G. Jarvis and R. B. Flavell, 1988. Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. *J. Mol. Biol.* **201:** 1-17.

Beckingham, K., 1981. The ribosomal DNA of *Calliphora erythrocephala*: the cistron classes of total genomic DNA. *J. Mol. Biol.* **149:** 141-169.

Begun, D. J., and C. F. Aquadro, 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356:** 519-520.

Belfort, M., 1993. An expanding universe of introns. *Science* **262:** 1009-1010.

Belkhiri, A., J. Buchko and G. R. Klassen, 1992. The 5S ribosomal RNA gene in *Pythium* species: two different genomic locations. *Mol. Biol. Evol.* **9:** 1089-1102.

Bhattacharya, D., B. Surek, M. Rüsing, S. Damberger and M. Melkonian, 1994. Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). *Proc. Natl. Acad. Sci. USA* **91:** 9916-9920.

Birky, C. J., Jr., and R. V. Skavaril, 1976. Maintenance of genetic homogeneity in systems with multiple genomes. *Genet. Res.* **27:** 249-265.

Bogenhagen, D. F., S. Sakonju and D. D. Brown, 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3' border of the region. *Cell* **19:** 27-35.

Botchan, P., R. H. Reeder and I. B. Dawid, 1977. Restriction analysis of the nontranscribed spacers of *Xenopus laevis* ribosomal DNA. *Cell* **11:** 599-607.

Brewer, B. J., D. Lockshon and W. L. Fangman, 1992. The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell* **71:** 267-276.

Britten, R. J., and D. E. Kohne, 1968. Repeated sequences in DNA. *Science* **161:** 529-540.

Broach, J. R., Y.-Y. Li, J. Feldman, M. Jayaram, J. Abraham, K. A. Nasmyth and J. B. Hicks, 1982. Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 1165-1173.

Brown, D. D., and K. Sugimoto, 1973. The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri. Cold Spring Harbor Symp. Quant. Biol.* **38:** 501-505.

Brown, D. D., P. C. Wensink and E. Jordan, 1972. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.* **63:** 57-73.

Brown, S. D. M., and G. A. Dover, 1981. Organization and evolutionary progress of a dispersed repetitive family of sequences in widely separated rodent genomes. *J. Mol. Biol.* **150:** 441-466.

Brownlee, A. G., 1988. A rapid DNA isolation procedure applicable to many refractory filamentous fungi. *Fungal Genet. Newsl.* **35:** 8-9.

Bryk, M., M. Banerjee, M. Murphy, K. E. Knudsen, D. J. Garfinkel and M. J. Curcio, 1997. Transcriptional silencing of Ty1 elements in the *RDN1* locus of yeast. *Genes Dev.* **11:** 255-269.

Bucheli, E., and A. Leutchmann, 1996. Evidence for genetic differentiation between choke-inducing and asymptomatic strains of the *Epichloë* grass endophyte from *Brachypodium sylvaticum. Evolution* **50:** 1879-1887.

Bullock, W. O., J. M. Fernandez and J. M. Short, 1987. XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5:** 376-378.

Buongiorno-Nardelli, M., F. Amaldi and P. A. Lava-Sanchez, 1972. Amplification as a retification mechanism for the redundant rRNA genes. *Nat. New Biol.* **238:** 134-137.

Butler, D. K., 1992. Ribosomal DNA is a site of chromosome breakage in aneuploid strains of *Neurospora. Genetics* **131:** 581-592.

Butler, D. K., and R. L. Metzenberg, 1989. Premeiotic change of nucleolus organizer size in *Neurospora*. *Genetics* **122:** 783-791.

Butler, D. K., and R. L. Metzenberg, 1990. Expansion and contraction of the nucleolus organizer region of *Neurospora*: changes originate in both proximal and distal segments. *Genetics* **126:** 325-333.

Callan, H. G., and L. Lloyd, 1960. Lampbrush chromosomes of crested newts *Triturus cristatus* (Laurenti). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **243:** 135-213.

Campbell, C. S., M. F. Wojciechowski, B. G. Baldwin, L. A. Alice and M. J. Donoghue, 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol. Biol. Evol.* **14:** 81-90.

Cannon, R. D., H. F. Jenkinson and M. G. Shepherd, 1990. Isolation and nucleotide sequence of an autonomously replicating sequence (ARS) element functional in *Candida albicans* and *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **221:** 210-218.

Carranza, S., G. Giribet, C. Ribera, J. Baguñà and M. Riutort, 1996. Evidence that two types of 18S rDNA coexist in the genome of *Dugesia (Schmidtea) mediterannea* (Platyhelminthes, Turbellaria, Tricladida). *Mol. Biol. Evol.* **13:** 824-832.

Cavalli, G., D. Bachmann and F. Thoma, 1996. Inactivation of topoisomerases affects transcription-dependent chromatin transitions in rDNA but not in a gene transcribed by RNA polymerase II. *EMBO J.* **15:** 590-597.

Cech, T. R., 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis - a review. *Gene* **73:** 259-271.

Cech, T. R., 1990. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59:** 543-568.

Charlesworth, B., C. H. Langley and W. Stephan, 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112:** 947-962.

Childs, G., R. Maxson, R. H. Cohn and L. Kedes, 1981. Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell* **23:** 651-663.

Cho, Y., Y.-L. Qiu, P. Kuhlman and J. D. Palmer, 1998. Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. USA* **95:** 14244-14249.

Christensen, M. J., A. Leuchtmann, D. D. Rowan and B. A. Tapper, 1993. Taxonomy of *Acremonium* endophytes of tall fescue (*Festuca arundinacea*), meadow fescue (*F. pratensis*) and perennial ryegrass (*Lolium perenne*). *Mycol. Res.* **97:** 1083-1092.

Christiansen, K., B. J. Bonven and O. Westergaard, 1987. Mapping of sequence-specific chromatin proteins by a novel method: Topoisomerase I on *Tetrahymena* ribosomal chromatin. *J. Mol. Biol.* **193:** 517-525.

Christman, M. F., F. S. Dietrich and G. R. Fink, 1988. Mitotic recombination in the rDNA of *S. cerevisiae* is suppressed by the combined action of DNA topoisomerases I and II. *Cell* **55:** 413-425.

Chung, K.-R., and C. L. Schardl, 1997. Vegetative compatibility between and within *Epichloë* species. *Mycologia* **89:** 558-565.

Clay, K., 1988 Clavicipitaceous fungal endophytes of grasses: coevolution and the change from parasitism to mutualism, pp. 79-105 in *Coevolution of Fungi with Plants and Animals*, edited by K. A. Pirozynski, and D. Hawksworth. Academic Press, London.

Clay, K., 1990. Fungal endophytes of grasses. *Annu. Rev. Ecol. Syst.* **21:** 275-297.

Cluster, P. D., D. Marinkovic, R. W. Allard and F. J. Ayala, 1987. Correlations between development rates, enzyme activities, ribosomal DNA spacer-length phenotypes, and adaption in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **84:** 610-614.

Coen, E. S., and G. A. Dover, 1983a. Multiple Pol I initiation sequences in rDNA spacers of *Drosophila melanogaster*. *Nucleic Acids Res.* **10:** 7017-7026.

Coen, E. S., and G. A. Dover, 1983b. Unequal exchanges and the coevolution of X and Y rDNA arrays in *Drosophila melanogaster*. *Cell* **33:** 849-855.

Coen, E. S., T. Strachan and G. A. Dover, 1982a. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Biol.* **158:** 17-35.

Coen, E. S., J. M. Thoday and G. Dover, 1982b. Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature* **295:** 564-568.

Collett, M. A., 1994 Cloning, characterisation and evolutionary relationships of two *pyr4* genes from an *Acremonium* endophyte of perennial ryegrass. PhD Thesis. Massey University, Palmerston North, New Zealand.

Collett, M. A., R. E. Bradshaw and D. B. Scott, 1995. A mutualistic fungal symbiont of perennial ryegrass contains two different *pyr4* genes, both expressing orotidine-5'-monophosphate decarboxylase. *Gene* **158:** 31-39.

Copenhaver, G. P., and C. S. Pikaard, 1996a. RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J.* **9:** 259-272.

Copenhaver, G. P., and C. S. Pikaard, 1996b. Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* **9:** 273-282.

Cox, M. M., 1997. Recombinational crossroads: eukaryotic enzymes and the limits of bacterial precedents. *Proc. Natl. Acad. Sci. USA* **94:** 11764-11766.

Crease, T. J., 1995. Ribosomal DNA evolution at the population level: Nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* **141:** 1327-1337.

Crease, T. J., and M. Lynch, 1991. Ribosomal DNA variation in *Daphnia pulex*. *Mol. Biol. Evol.* **8:** 620-640.

Cronn, R. C., X. Zhao, A. H. Paterson and J. F. Wendel, 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* **42:** 685-705.

Dammann, R., R. Lucchini, T. Koller and J. M. Sogo, 1995. Transcription in the yeast rRNA gene locus: distribution of the active gene copies and chromatin structure of their flanking regulatory sequences. *Mol. Cell. Biol.* **15:** 5294-5303.

Davies, R. W., R. B. Waring, J. A. Ray, T. A. Brown and C. Scazzocchio, 1982. Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300:** 719-724.

De Jonckheere, J. F., and S. Brown, 1998. Three different group I introns in the nuclear large subunit ribosomal DNA of the amoeboflagellate *Naegleria*. *Nucleic Acids Res.* **26:** 456-461.

Derr, L. K., and J. N. Strathern, 1993. A role for reverse transcripts in gene conversion. *Nature* **361:** 170-173.

Dobbs, D. L., W.-L. Shaiu and R. M. Benbow, 1994. Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucleic Acids Res.* **22:** 2479-2489.

Dobson, J. M., 1997 The molecular characterisation of the HMG CoA reductase gene from *Neotyphodium lolii.* M.Sc. Thesis. Massey University, New Zealand.

Dover, G. A., 1982. Molecular drive: a cohesive mode of species evolution. *Nature* **299:** 111-117.

Dover, G. A., 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* **2:** 159-165.

Dover, G. A., 1987. DNA turnover and the molecular clock. *J. Mol. Evol.* **26:** 47-58.

Dover, G. A., 1989. Linkage disequilibrium and molecular drive in the rDNA gene family. *Genetics* **122:** 249-252.

Dover, G. A., S. D. M. Brown, E. S. Coen, J. Dallas, T. Strachan and M. Trick, 1982 The dynamics of genome evolution and species differentiation, pp. 343-374 in *Genome Evolution*, edited by G. A. Dover, and R. B. Flavell. Academic Press, London.

Dover, G. A., A. R. Linares, T. Bowen and J. M. Hancock, 1993. Detection and quantification of concerted evolution and molecular drive. *Methods Enzymol.* **224:** 525-541.

Dower, W. J., J. F. Miller and C. W. Ragsdale, 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16:** 6127-6145.

Drouin, G., and M. Moniz de Sá, 1995. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol. Biol. Evol.* **12:** 481-493.

Dubcovsky, J., and J. Dvorák, 1995. Ribosomal RNA multigene loci: nomads of the Triticeae genomes. *Genetics* **140:** 1367-1377.

Dvorák, J., 1989 Evolution of multigene families: the ribosomal RNA gene loci of wheat and related species, pp. 83-97 in *Plant Populations Genetics, Breeding, and Genetic Resources*, edited by A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir. Sinauer Associates Inc., Sunderland, MA.

Dvorák, J., and R. Appels, 1982. Chromosome and nucleotide sequence differentiation in genomes of polyploid *Triticum* species. *Theor. Appl. Genet.* **63:** 349-360.

Dvorák, J., and R. Appels, 1986. Investigation of homologous crossing over and sister chromatid exchange in the wheat *Nor-B2* locus coding for rDNA and *Gli-B2* locus coding for gliadins. *Genetics* **113:** 1037-1056.

Dvorák, J., D. Jue and M. Lassner, 1987. Homogenization of tandemly repeated nucleotide sequences by distant-dependent nucleotide sequence conversion. *Genetics* **116:** 487-498.

Edelman, G. M., and J. A. Gally, 1970 Arrangement and evolution of eukaryotic genes, pp. 962-972 in *The Neurosciences: Second Study Program*, edited by F. O. Schmitt. Rockefeller University Press, New York.

Eggleston, A. K., and S. C. West, 1996. Exchanging partners: recombination in *E. coli*. *Trends Genet.* **12:** 20-26.

Eickbush, T. H., and W. D. Burke, 1985. Silkmoth chorion gene families contain patchwork patterns of sequence homology. *Proc. Natl. Acad. Sci. USA* **82:** 2814-2818.

Eickbush, T. H., and W. D. Burke, 1986. The silkmoth late chorion locus: II. Gradients of gene conversion in two paired multigene families. *J. Mol. Biol.* **190:** 357-366.

Elder, J. F., Jr., and B. J. Turner, 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70:** 297-320.

Ellis, R. E., J. E. Sulston and A. R. Coulson, 1986. The rDNA of *C. elegans*: Sequence and structure. *Nucleic Acids Res.* **14:** 2345-2364.

Endow, S. A., and K. C. Atwood, 1988. Magnification: Gene amplification by an inducible system of sister chromatid exchange. *Trends Genet.* **4:** 348-351.

Evans, D. H., and R. Kolodner, 1988. Effect of DNA structure and nucleotide sequence on Holliday junction resolution by a *Saccharomyces cerevisiae* endonuclease. *J. Mol. Biol.* **201:** 69-80.

Flavell, R. B., 1985 Repeated sequences and genome change, pp. 139-156 in *Genetic Flux in Plants*, edited by B. Hohn, and E. S. Dennis. Springer-Verlag, New York.

Flavell, R. B., and G. Martini, 1982 The genetic control of nucleolus formation with special reference to common breadwheat, pp. 113-128 in *The Nucleolus*, edited by E. G. Jordan, and C. A. Cullis. Cambridge University Press, Cambridge.

Fogel, S., J. W. Welch and E. J. Louis, 1984. Meiotic gene conversion mediates gene amplification in yeast. *Cold Spring Harbor Symp. Quant. Biol.* **49:** 55-65.

Frankham, R., 1988 Exchanges in the rRNA multigene family as a source of genetic variation, pp. 236-242 in *Proceedings of the Second International Conference on Quantitative Genetics*, edited by B. S. Weir, E. J. Eisen, M. M. Goodman and G. Namkoong. Sinauer Associates Inc., Sunderland, Mass.

Fukuoka, H., Y. Kageyama, K. Yamamoto and G. Takeda, 1994. Rapid conversion of rDNA intergenic spacer of diploid mutants of rice derived from γ-ray irradiated tetraploids. *Mol. Gen. Genet.* **243:** 166-172.

Gan, K. C., 1992 Isolation and characterization of a ribosomal RNA gene cluster from *Acremonium typhinum*. B.Sc. (Hons.) Thesis. Massey University, Palmerston North, New Zealand.

Gangloff, S., M. R. Lieber and R. Rothstein, 1994. Transcription, topoisomerases and recombination. *Experientia* **50:** 261-269.

Gangloff, S., H. Zou and R. Rothstein, 1996. Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO J.* **15:** 1715-1725.

Ganley, A. R. D., 1993 Extreme heterogeneity in the non-transcribed spacer region of Lp1 ribosomal DNA. B.Sc. (Hons.) Thesis. Massey University, New Zealand.

Garber, R. C., B. G. Turgeon, E. U. Selker and O. C. Yoder, 1988. Organization of ribosomal RNA genes in the fungus *Cochliobolus heterostrophus*. *Curr. Genet.* **14:** 573-582.

Garcia, S. N., and L. Pillus, 1999. Net results of nucleolar dynamics. *Cell* **97:** 825-828.

Geiduschek, E. P., and G. P. Tocchini-Valentini, 1988. Transcription by RNA polymerase III. *Annu. Rev. Biochem.* **57:** 873-914.

Gerbi, S., 1985 Evolution of ribosomal DNA, pp. 419-517 in *Molecular Evolutionary Genetics*, edited by R. J. MacIntyre. Plenum Press, New York.

Gerbi, S. A., 1986. The evolution of eukaryotic ribosomal DNA. *BioSystems* **19:** 247-258.

Gilbert, W., and D. Dressler, 1968. DNA replication: the rolling circle model. *Cold Spring Harbor Symp. Quant. Biol.* **33:** 473-484.

Gillings, M. R., R. Frankham, J. Speirs and M. Whalley, 1987. X-Y exchange and the coevolution of the X and Y rDNA arrays in *Drosophila melanogaster*. *Genetics* **116:** 241-251.

Glenn, A. E., C. W. Bacon, R. Price and R. Hanlin, 1996. Molecular phylogeny of *Acremonium* and its taxonomic implications. *Mycologia* **88:** 369-383.

Gottlieb, S., and R. E. Esposito, 1989. A new role for a yeast transcriptional silencer gene, *SIR2*, in regulation of recombination in ribosomal DNA. *Cell* **56:** 771-776.

Gottlieb, S., J. Wagstaff and R. E. Esposito, 1989. Evidence for two pathways of meiotic intrachromosomal recombination in yeast. *Proc. Natl. Acad. Sci. USA* **86:** 7072-7076.

Goyon, C., J.-L. Rossignol and G. Faugeron, 1996. Native DNA repeats and methylation in *Ascobolus*. *Nucleic Acids Res.* **24:** 3348-3356.

Graziani, F., R. Caizzi and S. Gargano, 1977. Circular ribosomal DNA during ribosomal magnification in *Drosophila melanogaster*. *J. Mol. Biol.* **112:** 49-63.

Gysler-Junker, A., Z. Bodi and J. Kohli, 1991. Isolation and characterization of *Schizosaccharomyces pombe* mutants affected in mitotic recombination. *Genetics* **128:** 495-504.

Haber, J. E., 1999. Gatekeepers of recombination. *Nature* **398:** 665-666.

Hadjiolov, A. A., 1984 *The Nucleolus and Ribosome Biogenesis*. Springer-Verlag Wien, New York.

Hawley, R. S., and C. H. Marcus, 1989. Recombinational controls of rDNA redundancy in *Drosophila*. *Annu. Rev. Genet.* **23:** 87-120.

Hawley, R. S., and K. D. Tartof, 1983. The ribosomal DNA of *Drosophila melanogaster* is organized differently from that of *Drosophila hydei*. *J. Mol. Biol.* **163:** 499-503.

Hibbett, D. S., 1996. Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. *Mol. Biol. Evol.* **13:** 903-917.

Hibner, B. L., W. D. Burke and T. H. Eickbush, 1991. Sequence identity in an early chorion multigene family is the result of localized gene conversion. *Genetics* **128:** 595-606.

Hijri, M., M. Hosny, D. van Tuinen and H. Dulieu, 1999. Intraspecific ITS polymorphism in *Scutellospora castanea* (Glomales, Zygomycota) is structured within multinucleate spores. *Fungal Genet. Biol.* **26:** 141-151.

Hillis, D. M., and M. T. Dixon, 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66:** 411-447.

Hillis, D. M., C. Moritz, C. A. Porter and R. J. Baker, 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **252:** 308-310.

Høgset, A., and T. B. Øyen, 1984. Correlation between suppressed meiotic recombination and the lack of DNA strand-breaks in the rRNA genes of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **12:** 7199-7213.

Holmes, D.S. and M. Quigley, 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114:** 193-197.

Holliday, R., 1964. A mechanism for gene conversion in fungi. *Genet. Res.* **5:** 282-304.

Hood, J. M., H. V. Huang and L. Hood, 1980. A computer simulation of evolutionary forces controlling the size of a multigene family. *J. Mol. Evol.* **15:** 181-196.

Hosny, M., M. Hijri, E. Passerieux and H. Dulieu, 1999. rDNA units are highly polymorphic in *Scutellospora castanea* (Glomales, Zygomycetes). *Gene* **226:** 61-71.

Hourcade, D., D. Dressler and J. Wolfson, 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. *Proc. Natl. Acad. Sci. USA* **70:** 2926-2930.

Ish-Horowitz, D., and J. F. Burke, 1981. Rapid and efficient cosmid cloning. *Nucleic Acids Res.* **9:** 2989-2998.

Israelewski, N., and E. R. Schmidt, 1982. Spacer size heterogeneity in ribosomal DNA of *Chironomus thummi* is due to a 120 bp repeat homologous to a predominantly centromeric repeated sequence. *Nucleic Acids Res.* **10:** 7689-7700.

Jagadeeswaran, P., B. G. Forget and S. M. Weissmann, 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? *Cell* **26:** 141-142.

Jakubczak, J. L., Y. Xiong and T. H. Eickbush, 1990. Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.* **212:** 37-52.

Jakubczak, J. L., M. K. Zenni, R. C. Woodruff and T. H. Eickbush, 1992. Turnover of R1 (Type I) and R2 (Type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. *Genetics* **131:** 129-142.

Jeffreys, A. J., A. MacLeod, K. Tamaki, D. L. Neil and D. G. Monckton, 1991. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354:** 204-209.

Jeffreys, A. J., R. Neumann and V. Wilson, 1990. Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60:** 473-485.

Jeffreys, A. J., K. Tamaki, A. MacLeod, D. G. Monckton, D. L. Neil and J. A. L. Armour, 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6:** 136-145.

Jinks-Robertson, S., M. Michelitch and S. Ramcharan, 1993. Substrate length requirements for efficient mitotic recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13:** 3937-3950.

Jinks-Robertson, S., and T. D. Petes, 1993. Experimental determination of rates of concerted evolution. *Methods Enzymol.* **224:** 631-646.

Johansen, S., T. Johansen and F. Haugli, 1992. Structure and evolution of myxomycete nuclear group I introns: a model for horizontal transfer by intron homing. *Curr. Genet.* **22:** 297-304.

Kaback, D. B., and H. O. Halvorson, 1977. Magnification of genes coding for ribosomal RNA in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **74:** 1177-1180.

Kamnert, I., L. Nielsen and J.-E. Edström, 1998. A concertedly evolving region in *Chironomus*, unique within the telomere. *J. Mol. Evol.* **46:** 562-570.

Kanaar, R., and J. H. J. Hoeijmakers, 1998. From competition to collaboration. *Nature* **391:** 335-338.

Keil, R. L., and A. D. McWilliams, 1993. A gene with specific and global effects on recombination of sequences from tandemly repeated genes in *Saccharomyces cerevisiae*. *Genetics* **135:** 711-718.

Keil, R. L., and G. S. Roeder, 1984. *Cis*-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. *Cell* **39:** 377-386.

Kellogg, E. A., and R. Appels, 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* **140:** 325-343.

Kim, R. A., and J. C. Wang, 1989. A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings. *Cell* **57:** 975-985.

Kimura, M., and T. Ohta, 1979. Population genetics of multigene family with special reference to decrease of genetic correlation with distance between gene members on a chromosome. *Proc. Natl. Acad. Sci. USA* **76:** 4001-4005.

Klein, H. L., 1995. Examination of mitotic recombination by means of hyper-recombination mutants in *Saccharomyces cerevisiae*. *Prog. Nucleic Acid Res. Mol. Biol.* **51:** 271-303.

Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31:** 147-157.

Krystal, M., and N. Arnheim, 1978. Length heterogeneity in a region of the human ribosomal gene spacer is not accompanied by extensive population polymorphism. *J. Mol. Biol.* **126:** 91-104.

Krystal, M., P. D'Eustachio, F. H. Ruddle and N. Arnheim, 1981. Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants. *Proc. Natl. Acad. Sci. USA* **78:** 5744-5748.

Kulkens, T., C. A. F. M. van der Sande, A. F. Dekker, H. van Heerikhuizen and R. J. Planta, 1992. A system to study transcription by yeast RNA polymerase I within the chromosomal context: functional analysis of the ribosomal DNA enhancer and the RBP1/REB1 binding sites. *EMBO J.* **11:** 4665-4674.

Lambowitz, A. M., 1989. Infectious introns. *Cell* **56:** 323-326.

Lambowitz, A. M., and M. Belfort, 1993. Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62:** 587-622.

Lassner, M., O. Anderson and J. Dvorák, 1987. Hypervariation associated with a 12-nucleotide direct repeat and inferences on intergenomic homogenization of ribosomal RNA gene spacers based on the DNA sequence of a clone from the wheat *Nor-D3* locus. *Genome* **29:** 770-781.

Lassner, M., and J. Dvorák, 1986. Preferential homogenization between adjacent and alternate subrepeats in wheat rDNA. *Nucleic Acids Res.* **14:** 5499-5512.

Lathe, W. C., III, and T. H. Eickbush, 1997. A single lineage of R2 retrotransposable elements is an active, evolutionarily stable component of the *Drosophila* rDNA locus. *Mol. Biol. Evol.* **14:** 1232-1241.

Li, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland.

Liao, D., 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64:** 24-30.

Liao, D., T. Pavelitz, J. R. Kidd, K. K. Kidd and A. M. Weiner, 1997. Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J.* **16:** 588-598.

Lichten, M., and A. S. H. Goldman, 1995. Meiotic recombination hotspots. *Annu. Rev. Genet.* **29:** 423-444.

Linares, A. R., T. Bowen and G. A. Dover, 1994. Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster*. *J. Mol. Evol.* **39:** 151-159.

Liu, Y., M. Rocourt, S. Pan, C. Liu and M. J. Leibowitz, 1992. Sequence and variability of the 5.8S and 26S rRNA genes of *Pneumocystis carinii*. *Nucleic Acids Res.* **20:** 3763-3772.

Lohe, A. R., and P. A. Roberts, 1990. An unusual Y chromosome of *Drosophila simulans* carrying amplified rDNA spacer without rRNA genes. *Genetics* **125:** 399-406.

Loidl, J., and K. Nairz, 1997. Karyotype variability in yeast caused by nonallelic recombination in haploid meiosis. *Genetics* **146:** 79-88.

Long, E. O., and I. B. Dawid, 1980. Repeated genes in eukaryotes. *Annu. Rev. Biochem.* **49:** 727-764.

Lyckegaard, E. M. S., and A. G. Clark, 1991. Evolution of ribosomal RNA gene copy number on the sex chromosomes of *Drosophila melanogaster*. *Mol. Biol. Evol.* **8:** 458-474.

Lynch, M., R. Bürger, D. Butcher and W. Gabriel, 1993. The mutational meltdown in asexual populations. *J. Hered.* **84:** 339-344.

MacGregor, H. C., 1982 Ways of amplifying ribosomal genes, pp. 129-151 in *The Nucleolus*, edited by E. G. Jordan, and C. A. Cullis. Cambridge University Press, Cambridge.

Mao, J., B. Appel, J. Schaack, S. Sharp, H. Yamada and D. Söll, 1982. The 5S RNA genes of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **10:** 487-500.

Martin, F. N., 1990. Variation in the ribosomal DNA repeat unit within single-oospore isolates of the genus Pythium. *Genome* **33:** 585-591.

Mercure, S., S. Montplaisir and G. Lemay, 1993. Correlation between the presence of a self-splicing intron in the 25S rDNA of *C. albicans* and strains susceptibility to 5-fluorocytosine. *Nucleic Acids Res.* **21:** 6020-6027.

Meselson, M. S., and C. M. Radding, 1975. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* **72:** 358-361.

Metzenberg, R. L., J. N. Stevens, E. U. Selker and E. Morzycka-Wroblewska, 1985. Identification and chromosomal distribution of 5S rRNA genes in *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **82:** 2067-2071.

Mian, A., and G. A. Dover, 1990. Promoter variation in the ribosomal RNA genes in *Drosophila melanogaster* strains. *Nucleic Acids Res.* **18:** 3795-3801.

Michel, F., and B. F. Lang, 1985. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature* **316:** 641-643.

Mohr, G., and A. M. Lambowitz, 1991. Integration of a group I intron into a ribosomal RNA sequence promoted by a tyrosyl-tRNA synthetase. *Nature* **354:** 164-167.

Moon, C. D., B. A. Tapper and B. Scott, 1999. Identification of *Epichloë* endophytes *in planta* by a microsatellite-based PCR fingerprinting assay with automated analysis. *Appl. Env. Microbiol.* **65:** 1268-1279.

Morton, D. G., and K. U. Sprague, 1984. *In vitro* transcription of a silkworm 5S RNA gene requires an upstream signal. *Proc. Natl. Acad. Sci. USA* **81:** 5519-5522.

Morzycka-Wroblewska, E., E. U. Selker, J. N. Stevens and R. L. Metzenberg, 1985. Concerted evolution of dispersed *Neurospora crassa* 5S RNA genes: pattern of sequence conservation between allelic and nonallelic genes. *Mol. Cell. Biol.* **5:** 46-51.

Moss, T., K. Mitchelson and R. De Winter, 1985. The promotion of ribosomal transcription in eukaryotes. *Oxf. Surv. on Eucaryot. Genes* **2:** 207-250.

Mota, E. M., and R. A. Collins, 1988. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature* **332:** 654-656.

Muller, H. J., 1964. The relation of recombination to mutational advance. *Mutat. Res.* **1:** 2-9.

Munz, P., H. Amstutz, J. Kohli and U. Leupold, 1982. Recombination between dispersed serine tRNA genes in *Schizosaccharomyces pombe*. *Nature* **300:** 225-231.

Murray, F. R., G. C. M. Latch and D. B. Scott, 1992. Surrogate transformation of perennial ryegrass, *Lolium perenne*, using genetically modified *Acremonium* endophyte. *Mol. Gen. Genet.* **233:** 1-9.

Muscarella, D. E., and V. M. Vogt, 1989. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **56:** 443-454.

Muscarella, D. E., and V. M. Vogt, 1993. A mobile group I intron from *Physarum polycephalum* can insert itself and induce point mutations in the nuclear ribosomal DNA of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13:** 1023-1033.

Muscarella, D. E., V. M. Vogt and S. E. Bloom, 1987. Characterization of ribosomal RNA synthesis in a gene dosage mutant: the relationship of topoisomerase I and chromatin structure to transcriptional activity. *J. Cell Biol.* **105:** 1501-1513.

Nagylaki, T., 1984. Evolution of multigene families under interchromosomal gene conversion. *Proc. Natl. Acad. Sci. USA* **81:** 3796-3800.

Nagylaki, T., and T. D. Petes, 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100:** 315-337.

Neuvéglise, C., Y. Brygoo and G. Riba, 1997. 28S rDNA group-I introns: a powerful tool for identifying strains of *Beauveria brongniartii*. *Mol. Ecol.* **6:** 373-381.

Nishida, H., and J. Sugiyama, 1995. A common group I intron between a plant parasitic fungus and its host. *Mol. Biol. Evol.* **12:** 883-886.

O'Donnell, K., and E. Cigelnik, 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* **7:** 103-116.

O'Kane, S. L., Jr., B. A. Schaal and I. A. Al-Shehbaz, 1996. The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. *Syst. Bot.* **21:** 559-566.

Odorico, D. M., and D. J. Miller, 1997. Variation in the ribosomal internal transcribed spacers and 5.8S rDNA among five species of *Acropora* (Cnidaria; Scleractinia): patterns of variation consistent with reticulate evolution. *Mol. Biol. Evol.* **14:** 465-473.

Ohta, T., 1976. Simple model for treating evolution of multigene families. *Nature* **263:** 74-76.

Ohta, T., 1977. On the gene conversion model as a mechanism for maintenance of homogeneity in systems with multiple genomes. *Genet. Res.* **30:** 89-91.

Ohta, T., 1983. On the evolution of multigene families. *Theor. Popul. Biol.* **23:** 216-240.

Ohta, T., and G. A. Dover, 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc. Natl. Acad. Sci. USA* **80:** 4079-4083.

Ohta, T., and G. A. Dover, 1984. The cohesive population genetics of molecular drive. *Genetics* **108:** 501-521.

Orr-Weaver, T. L., and J. W. Szostak, 1985. Fungal recombination. *Microbiol. Rev.* **49:** 33-58.

Ozenberger, B. A., and G. S. Roeder, 1991. A unique pathway of double-strand break repair operates in tandemly repeated genes. *Mol. Cell. Biol.* **11:** 1222-1231.

Pavelitz, T., L. Rusché, A. G. Matera, J. M. Scharf and A. M. Weiner, 1995. Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ,* without changing the cytological context of the *RNU2* locus. *EMBO J.* **14:** 169-177.

Pellegrini, M., J. Manning and N. Davidson, 1977. Sequence arrangement of the rDNA of *Drosophila melanogaster. Cell* **10:** 213-224.

Peng, M., P. A. Lemke and N. K. Singh, 1993. A nucleotide sequence involved in replicative transformation of a filamentous fungus. *Curr. Genet.* **24:** 114-121.

Perelson, A. S., and G. I. Bell, 1977. Mathematical models for the evolution of multigene families by unequal crossing over. *Nature* **265:** 304-310.

Petes, T. D., 1980. Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell* **19:** 765-774.

Petes, T. D., and D. Botstein, 1977. Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. *Proc. Natl. Acad. Sci. USA* **74:** 5091-5095.

Petes, T. D., and C. W. Hill, 1988. Recombination between repeated genes in microorganisms. *Annu. Rev. Genet.* **22:** 147-168.

Pieler, T., J. Hamm and R. G. Roeder, 1987. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. *Cell* **41:** 91-100.

Pingoud, A., C. Urbanke, J. Alves, H.-J. Ehbrecht, M. Zabeau and C. Gualerzi, 1984. Effect of polyamines and basic proteins on cleavage of DNA by restriction endonucleases. *Biochemistry* **23:** 5697-5703.

Polanco, C., A. I. González, A. de la Fuente and G. A. Dover, 1998. Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. *Genetics* **149:** 243-256.

Polans, N. O., N. F. Weeden and W. F. Thompson, 1986. Distribution, inheritance and linkage relationships of ribosomal DNA spacer length variants in pea. *Theor. Appl. Genet.* **72:** 289-295.

Pont, G., F. Degroote and G. Picard, 1987. Some extrachromosomal circular DNAs from *Drosophila* embryos are homologous to tandemly repeated genes. *J. Mol. Biol.* **195:** 447-451.

Pukkila, P. J., and C. Skrzyniä, 1993. Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete *Coprinus cinereus*. *Genetics* **133:** 203-211.

Rasooly, R. S., and L. G. Robbins, 1991. *Rex* and a suppressor of *Rex* are repeated neomorphic loci in the *Drosophila melanogaster* ribosomal DNA. *Genetics* **129:** 119-132.

Reddy, P., and R. Appels, 1989. A second locus for the 5S multigene family in *Secale* L.: sequence divergence in two lineages of the family. *Genome* **32:** 456-467.

Reddy, R., D. Henning, L. Rothblum and H. Busch, 1986. Some gene variants for 5S RNA are dispersed in the rat genome. *J. Biol. Chem.* **261:** 10618-10623.

Reeder, R. H., D. D. Brown, P. K. Wellauer and I. B. Dawid, 1976. Patterns of ribosomal DNA spacer lengths are inherited. *J. Mol. Biol.* **105:** 507-516.

Reeder, R. H., and J. G. Roan, 1984. The mechanism of nucleolar dominance in *Xenopus* hybrids. *Cell* **38:** 38-44.

Reeder, R. H., J. G. Roan and M. Dunaway, 1983. Spacer regulation of *Xenopus* ribosomal gene transcription: Competition in oocytes. *Cell* **35:** 449-456.

Renkawitz-Pohl, R., K. H. Glätzer and W. Kunz, 1981. Ribosomal RNA genes with an intervening sequence are clustered within the X chromosomal ribosomal DNA of *Drosophila hydei*. *J. Mol. Biol.* **148:** 95-101.

Rieseberg, L. H., 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am. J. Bot.* **78:** 1218-1237.

Ritossa, F., F. Scalenghe, N. Di Turi and A. M. Contini, 1973. On the cell stage of X-Y recombination during rDNA magnification in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 483-490.

Rochaix, J.-D., and A. P. Bird, 1975. Circular ribosomal DNA and ribosomal DNA: replication in somatic amphibian cells. *Chromosoma* **52:** 317-327.

Rockmill, B., J. Engebrecht, H. Scherthan, J. Loidl and G. S. Roeder, 1995. The yeast *MER2* gene is required for chromosome synapsis and the initiation of meiotic recombination. *Genetics* **141:** 49-59.

Rodland, K. D., and P. J. Russell, 1982. Regulation of ribosomal RNA cistron number in a strain of *Neurospora crassa* with a duplication of the nucleolus organizer region. *Biochim. Biophys. Acta* **697:** 162-169.

Rogers, S. O., and A. J. Bendich, 1987a. Heritability and variability in ribosomal RNA genes of *Vicia faba*. *Genetics* **117:** 285-295.

Rogers, S. O., and A. J. Bendich, 1987b. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* **9:** 509-520.

Russell, P. J., R. C. Petersen and S. Wagner, 1988. Ribosomal DNA inheritance and recombination in *Neurospora crassa. Mol. Gen. Genet.* **211:** 541-544.

Russell, P. J., and K. D. Rodland, 1986. Magnification of rRNA gene number in a *Neurospora crassa* strain with a partial deletion of the nucleolus organizer. *Chromosoma* **93:** 337-340.

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen and R. W. Allard, 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc. Natl. Acad. Sci. USA* **81:** 8014-8018.

Sakonju, S., D. F. Bogenhagen and D. D. Brown, 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region. *Cell* **19:** 13-25.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.

Samson, M.-L., and M. Wegnez, 1988. Bipartite structure of the 5S ribosomal gene family in a *Drosophila melanogaster* strain, and its evolutionary implications. *Genetics* **118:** 685-691.

Sang, T., D. J. Crawford and T. F. Stuessy, 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. USA* **92:** 6813-6817.

Sastri, D. C., K. Hilu, R. Appels, E. S. Lagudah, J. Playford and B. R. Baum, 1992. An overview of evolution in plant 5S DNA. *Plant Syst. Evol.* **183:** 169-181.

Schäfer, M., A. R. Wyman and R. White, 1981. Length variation in the non-transcribed spacer of *Calliphora erythrocephala* ribosomal DNA is due to a 350 base-pair repeat. *J. Mol. Biol.* **146:** 179-199.

Schardl, C. L., 1996. *Epichloë* species: Fungal symbionts of grasses. *Annu. Rev. Phytopathol.* **34:** 109-130.

Schardl, C. L., A. Leuchtmann, H.-F. Tsai, M. A. Collett, D. M. Watt and D. B. Scott, 1994. Origin of a fungal symbiont of perennial ryegrass by interspecific hybridisation of a mutualist with the ryegrass choke pathogen, *Epichloë typhina. Genetics* **136:** 1307-1317.

Schardl, C. L., J.-S. Liu, J. F. White, Jr., R. A. Finkel, Z. An and M. R. Siegel, 1991. Molecular phylogenetic relationships of nonpathogenic grass mycosymbionts and clavicipitaceous plant pathogens. *Plant Syst. Evol.* **178:** 27-41.

Schlötterer, C., and D. Tautz, 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* **4:** 777-783.

Scott, B., and C. Schardl, 1993. Fungal symbionts of grasses: evolutionary insights and agricultural potential. *TIM* **1:** 196-200.

Sekiguchi, J., N. C. Seeman and S. Shuman, 1996. Resolution of Holliday junctions by eukaryotic DNA topoisomerase I. *Proc. Natl. Acad. Sci. USA* **93:** 785-789.

Selker, E. U., E. Morzycka-Wroblewska, J. N. Stevens and R. L. Metzenberg, 1986. An upstream signal is required for *in vitro* transcription of *Neurospora* 5S RNA genes. *Mol. Gen. Genet.* **205:** 189-192.

Selker, E. U., C. Yanofsky, K. Driftmier, R. L. Metzenberg, B. Alzner-DeWeerd and U. L. RajBhandary, 1981. Dispersed 5S RNA genes in *N. crassa*: structure, expression and evolution. *Cell* **24:** 819-828.

Shah, R., R. J. Bennett and S. C. West, 1994. Genetic recombination in *E. coli* : RuvC protein cleaves Holliday junctions at resolution hotspots *in vitro*. *Cell* **79:** 853-864.

Sharp, Z. D., V. V. Gandhi and J. D. Procunier, 1983. X chromosome nucleolus organizer mutants which alter major type I repeat multiplicity in *Drosophila melanogaster*. *Mol. Gen. Genet.* **190:** 438-443.

Shen, W.-H., and B. Hohn, 1992. DMSO improves PCR amplification of DNA with complex secondary structure. *Trends Genet.* **8:** 227.

Shufran, K. A., D. C. Peters and J. A. Webster, 1997. Generation of clonal diversity by sexual reproduction in the greenbug, *Schizaphis graminum*. *Insect Mol. Biol.* **6:** 203-209.

Siegal, M. R., and C. L. Schardl, 1991 Fungal endophytes of grasses: detrimental and beneficial associations, pp. 198-221 in *Microbial Ecology of Leaves*, edited by J. H. Andrew, and S. S. Hirano. Springer Verlag, Berlin.

Sinclair, D. A., and L. Guarente, 1997. Extrachromosomal rDNA circles - a cause of aging in yeast. *Cell* **91:** 1033-1042.

Singer, M., and P. Berg, 1991 *Genes and Genomes.* University Science Books, Mill Valley.

Sjamsuridzal, W., and J. Sugiyama, 1998. Detection of multiple insertions of group I introns in the nuclear small subunit ribosomal RNA gene from the ustilaginomycete *Graphiola phoenicis* parasitic on palm. *J. Gen. Appl. Microbiol.* **44:** 355-360.

Smith, G. P., 1973. Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 507-513.

Smith, G. P., 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* **191:** 528-535.

Smith, G. R., 1994. Hotspots of homologous recombination. *Experientia* **50:** 234-241.

Smith, J. S., C. B. Brachmann, L. Pillus and J. D. Boeke, 1998. Distribution of a limited Sir2 protein pool regulates the strength of yeast rDNA silencing and is modulated by Sir4p. *Genetics* **149:** 1205-1219.

Sogin, M. L., A. Ingold, M. Karlok, H. Nielsen and J. Enberg, 1986. Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major *Tetrahymena* groups. *EMBO J.* **5:** 3625-3630.

Soltis, P. S., and D. E. Soltis, 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. *Syst. Bot.* **16:** 407-413.

Southern, E. M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503-517.

Stephan, W., and S. Cho, 1994. Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* **136:** 333-341.

Stewart, S. E., and G. S. Roeder, 1989. Transcription by RNA polymerase I stimulates mitotic recombination in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **9:** 3464-3472.

Strachan, T., D. Webb and G. A. Dover, 1985. Transition stages of molecular drive in multiple-copy DNA families in *Drosophila. EMBO J.* **4:** 1701-1708.

Sugawara, N., and J. E. Haber, 1992. Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Mol. Cell. Biol.* **12:** 563-575.

Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein and F. W. Stahl, 1983. The double-strand-break repair model for recombination. *Cell* **33:** 25-35.

Szostak, J. W., and R. Wu, 1980. Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae. Nature* **284:** 426-430.

Tabata, S., 1981. Nucleotide sequences of the 5S ribosomal RNA genes and their adjacent regions in *Schizosaccharomyces pombe. Nucleic Acids Res.* **9:** 6429-6437.

Tan, M. K., 1997. Origin and inheritance of group I introns in 26S rRNA genes of *Gaeumannomyces graminis. J. Mol. Evol.* **44:** 637-645.

Tartof, K. D., 1973. Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 491-500.

Tautz, D., C. Tautz, D. Webb and G. A. Dover, 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. *J. Mol. Biol.* **195:** 525-542.

Terracol, R., 1987. Differential magnification of rDNA gene types in bobbed mutants of *Drosophila melanogaster. Mol. Gen. Genet.* **208:** 168-176.

Thomas, B. J., and R. Rothstein, 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* **56:** 619-630.

Thompson, A. J., and D. L. Herrin, 1994. A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8S rRNA. Implications for intron-mediated RNA recombination, intron transposition and 5.8S rRNA structure. *J. Mol. Biol.* **236:** 455-468.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673-4680.

Thompson, W. F., and R. B. Flavell, 1988. DNaseI sensitivity of ribosomal RNA genes in chromatin and nucleolar dominance in wheat. *J. Mol. Biol.* **204:** 535-548.

Tsai, H.-F., J.-S. Liu, C. Staben, M. J. Christensen, G. C. M. Latch, M. R. Siegel and C. L. Schardl, 1994. Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloë* species. *Proc. Natl. Acad. Sci. USA* **91:** 2542-2546.

Tulin, A. V., G. L. Kogan, D. Filipp, M. D. Balakireva and V. A. Gvozdev, 1997. Heterochromatic *Stellate* gene cluster in *Drosophila melanogaster* : Structure and molecular evolution. *Genetics* **146:** 253-262.

Turmel, M., R. R. Gutell, J.-P. Mercier, C. Otis and C. Lemieux, 1993. Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa. Three internal transcribed spacers and 12 group I intron insertion sites. *J. Mol. Biol.* **232:** 446-467.

Van Heerikhuizen, H., A. Ykema, J. Klootwijk, C. Gaillardin, C. Ballas and P. Fournier, 1985. Heterogeneity in the ribosomal RNA genes of the yeast *Yarrowia lipolytica*; cloning and analysis of two size classes of repeats. *Gene* **39:** 213-222.

van Houten, W. H. J., N. Scarlett and K. Bachmann, 1993. Nuclear DNA markers of the Australian tetraploid *Microseris scapigera* and its North American diploid relatives. *Theor. Appl. Genet.* **87:** 498-505.

Van Oppen, M. J. H., J. L. Olsen and W. T. Stam, 1993. Evidence for independent acquisition of group I introns in green algae. *Mol. Biol. Evol.* **10:** 1317-1326.

Vaughan, H. E., M. Jamilena, C. Ruiz Rejón, J. S. Parker and M. A. Garrido-Ramos, 1993. Loss of nucleolar-organizer regions during polyploid evolution in *Scilla autumnalis*. *Heredity* **71:** 574-580.

Viaud, M., Y. Couteaudier, C. Levis and G. Riba, 1996. Genome organization in *Beauveria bassiana*: electrophoretic karyotype, gene mapping, and telomeric fingerprint. *Fungal Genet. Biol.* **20:** 175-183.

Vieira, J., and J. Messing, 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153:** 3-11.

Voelkel-Meiman, K., R. L. Keil and G. S. Roeder, 1987. Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell* **48:** 1071-1079.

Vogler, A. P., and R. DeSalle, 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. *Mol. Biol. Evol.* **11:** 393-405.

Walsh, J. B., 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics* **115:** 553-567.

Waters, E. R., and B. A. Schaal, 1996. Biased gene conversion is not occurring among rDNA repeats in the *Brassica* triangle. *Genome* **39:** 150-154.

Waugh O'Neill, R. J., M. J. O'Neill and J. A. Marshall Graves, 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393:** 68-72.

Welch, J. W., D. H. Maloney and S. Fogel, 1990. Unequal crossing-over and gene conversion at the amplified *CUP1* locus of yeast. *Mol. Gen. Genet.* **222:** 304-310.

Wellauer, P. K., and I. B. Dawid, 1978. Ribosomal DNA in *Drosophila melanogaster*: II. Heteroduplex mapping of cloned and uncloned rDNA. *J. Mol. Biol.* **126:** 769-782.

Wellauer, P. K., I. B. Dawid, D. D. Brown and R. H. Reeder, 1976a. The molecular basis for length heterogeneity in ribosomal DNA from *Xenopus laevis*. *J. Mol. Biol.* **105:** 461-486.

Wellauer, P. K., R. H. Reeder, I. B. Dawid and D. D. Brown, 1976b. The arrangement of length heterogeneity in repeating units of amplified and chromosomal ribosomal DNA from *Xenopus laevis*. *J. Mol. Biol.* **105:** 487-505.

Wendel, J. F., A. Schnabel and T. Seelanan, 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **92:** 280-284.

White, T. J., T. Bruns, S. Lee and J. Taylor, 1990 Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, pp. 315-322 in *PCR Protocols: A Guide to Methods and Applications*, edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White. Academic Press, San Diego.

Whitehouse, H. L. K., 1982 *Genetic Recombination, Understanding the Mechanisms*. John Wiley and Sons, New York.

Williams, S. M., 1990. The opportunity for natural selection on multigene families. *Genetics* **124:** 439-441.

Williams, S. M., R. DeSalle and C. Strobeck, 1985. Homogenization of geographical variants at the nontranscribed spacer of rDNA in *Drosophila mercatorum*. *Mol. Biol. Evol.* **2:** 338-346.

Williams, S. M., J. A. Kennison, L. G. Robbins and C. Strobeck, 1989. Reciprocal recombination and the evolution of the ribosomal gene family of *Drosophila melanogaster*. *Genetics* **122:** 617-624.

Williams, S. M., L. G. Robbins, P. D. Cluster, R. W. Allard and C. Strobeck, 1990. Superstructure of the *Drosophila* ribosomal gene family. *Proc. Natl. Acad. Sci. USA* **87:** 3156-3160.

Williams, S. M., and C. Strobeck, 1985. Sister chromatid exchange and the evolution of rDNA spacer length. *J. Theor. Biol.* **116:** 625-636.

Woodson, S. A., and T. R. Cech, 1989. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. *Cell* **57:** 335-345.

Worton, R. G., J. Sutherland, J. E. Sylvester, H. F. Willard, S. Bodrug, I. Dubé, C. Duff, V. Kean, P. N. Ray and R. D. Schmickel, 1988. Human ribosomal RNA genes: orientation of the tandem array and conservation of the 5' end. *Science* **239:** 64-68.

Wu, T.-C., and M. Lichten, 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* **263:** 515-518.

Wu, T.-C., and M. Lichten, 1995. Factors that affect the location and frequency of meiosis-induced double-strand breaks in *Saccharomyces cerevisiae*. *Genetics* **140:** 55-66.

Yuan, L.-W., and R. L. Keil, 1990. Distance-independence of mitotic intrachromosomal recombination in *Saccharomyces cerevisiae*. *Genetics* **124:** 263-273.

Zamb, T. J., and T. D. Petes, 1982. Analysis of the junction between ribosomal RNA genes and single-copy chromosomal sequences in the yeast *Saccharomyces cerevisiae*. *Cell* **28:** 355-364.

Zhang, D., and T. Sang, 1999. Physical mapping of ribosomal RNA genes in peonies (*Paeonia*, Paeoniaceae) by fluorescent *in situ* hybridisation: implications for phylogeny and concerted evolution. *Am. J. Bot.* **86:** 735-740.

Zijlstra, C., A. E. M. Lever, B. J. Uenk and C. H. Van Silfhout, 1995. Differences between ITS regions of isolates of root-knot nematodes *Meloidogyne hapla* and *M. chitwoodi*. *Phytopathology* **85:** 1231-1237.

Zimmer, E. A., S. L. Martin, S. M. Beverley, Y. W. Kan and A. C. Wilson, 1980. Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* **77:** 2158-2162.

Zwick, M. G., M. Wiggs and M. R. Paule, 1991. Sequence and organization of 5S RNA genes from the eukaryotic protist *Acanthamoeba castellani*. *Gene* **101:** 153-157.

# CORRECTIONS

## CHAPTER ONE

| | | |
|---|---|---|
| Page 2 | line 9 | replace "Siegal" with "Siegel" |

## CHAPTER TWO

| | | |
|---|---|---|
| Page 14 | line 14 | replace "mycelia" with "mycelium" |
| Page 27 | line 15 | insert "Size markers were detected radiographically by radiolabelling the appropriate size marker as set out in Section 2.10.2. The radiolabelled size marker was then included in the hybridisation (Section 2.10.3)." |
| Page 32 | lines 1 and 3 | replace "automatic" with "automated" |

## CHAPTER THREE

| | | |
|---|---|---|
| Page 41 | line 18 | The four laboratory cultures are indicated by the first letter following the Lp1 designation along with a zero following the first letter (e.g. Lp1A0). The first-round single-spore isolates are designated by an integer following the laboratory culture letter designation (e.g. Lp1A3). |
| Page 44 | line 7 | replace "indicated by arrows." to "indicated by arrows in panels C, D and E." |
| Page 55 | line 18 | replace "homology" with "similarity" |
| Page 65 | line 19 | replace "was very weak, presumably .... in the IGS sequence" with "was very weak. This weak hybridisation is unlikely be the result of sequence divergence. The probe used derives from Lp1, and therefore is effectively from E8 (see Section 3.3.1). This probe hybridises strongly to the Lp5 IGS despite Lp5 and E8 being distantly-related endophytes. The weak hybridisation is not the result of inconsistent gel loading, but its cause remains unknown. |
| Page 66 | lines 5 and 6 | replace "*coenophilum*" with "*coenophialum*" |
| Page 74 | line 7 | replace "thin lines and the" with "thin lines and exon" |
| Page 77 | line 29 | The Southern blot was probed with a probe made from the nts1 - nts2 PCR product, not the 5.6 kb *Sal*I coding region probe. |
| Page 77 | line 37 | insert "Digestion of the chromosomes with I-*Ppo*I does not produce drastic changes in the chromosomal banding pattern, although some chromosomal bands are observed to change size." |
| Page 78 | lines 8 & 12 | insert "The positions of the wells are indicated." |
| Page 80 | lines 9 & 13 | insert "The positions of the wells are indicated." |
| Page 82 | lines 7 & 10 | insert "The positions of the wells are indicated." |

| | | |
|---|---|---|
| Page 83 | lines 9 & 12 | insert "The positions of the wells are indicated." |
| Page 84 | line 37 | replace "an rDNA primer adjacent" with "an rDNA primer that binds to the sequence adjacent" |
| Page 91 | line 15 | replace "signal was low" with "signal is low" |

CHAPTER FOUR

| | | |
|---|---|---|
| Page 99 | line 38 | replace "unequal crossing as" with "unequal crossing over as" |
| Page 103 | line 7 | replace "(Jinks-Robertson *et al.*, 1993; Yuan and Keil, 1990) found that" with "Others (Jinks-Robertson *et al.*, 1993; Yuan and Keil, 1990) have found that" |
| Page 104 | line 11 | replace "that of prokaryote, but" with "that of prokaryote recombination, but" |
| Page 111 | line 21 | nuclear "pickup" describes the event where a trichogyne from an ascogonium undergoes anastomosis with a hypha from another, compatible fungus and one or more nuclei migrate into the ascogonium. |
| Page 118 | line 1 | insert "The weak hybridisation of the IGS probe with the genomic DNA from these other hybrids makes it difficult to draw firm conclusions on the nature of the IGS in these hybrids. It remains possible that IGS length heterogeneity is a feature of many or all of these hybrids." |
| Page 122 | line 34 | replace "clearly demonstrates" with "indicates" |
| Page 123 | line 11 | insert "It is interesting to note that digestion of the chromosomes with the ribosomal-specific endonuclease, I-*Ppo*I, does not produce drastic changes in the sizes of the chromosomal bands for Lp1 or Lp5. Such a result might be expected if the rDNA arrays are located telomerically. However the complexity of the chromosomal banding patterns for these endophytes makes it difficult to determine the fates of the chromosomal bands that contain rDNA arrays." |
| Page 123 | line 33 | insert "For instance, if the rDNA arrays are located telomerically and the telomeres are not replaced in older cells (due to failure of the telomerase), then degradation of the rDNA could occur in these older cells, giving a smear of rDNA array sizes." |
| Page 127 | line 29 | replace "speciation of Lp5" with "speciation of *N. lolii*" |
| Page 132 | line 31 | replace "basis" with "bases" |

BIBLIOGRAPHY

| | | |
|---|---|---|
| Page 161 | line 36 | replace "Siegal" with "Siegel" |