A MOLECULAR ANALYSIS OF FUNGAL MATING TYPE GENES

Julie Nixon

A thesis presented for the degree of PhD
University of Edinburgh
1997

Declaration

This thesis and all the work herein was composed by myself, unless otherwise stated.

Julie Nixon

September 1997.

Acknowledgements

I would like to thank Jeff Bond for advice, help, encouragement and patience and David Finnegan and all those in the Finnegan lab, past and present for taking in a stray. Thanks also to Frank Wright for his expert help with phylogenetic analysis. I am grateful to all on the 5th and 6th floors for the friendships they provided.

Thanks also to my friends outside of work without whom I would not have preserved my sanity. In particular I would like to thank those at Murrayfield ice-rink with whom I spend many happy hours accumulating bruises in the name of relaxation.

Finally many thanks to my family and Alastair for their love and continuing support.

ABBREVIATIONS

A	-adenosine
amp	-ampicillin
bp	-base pair
C	-cytidine
°C	-degrees Celsius
CTAB	-Hexadecyltrimethylammonium bromide
dd	-dideoxy
DMSO	-dimethyl sulphoxide
DNase	-deoxyribonuclease
dNTP	-deoxyribonucleoside triphosphate
DTT	-dithiothreitol
EDTA	-diaminiethanetetra-acetic acid
g	-gram
G	-guanosine
IPTG	-isopropyl-β-D-thiogalactopyranoside
kb (s)	-kilobase (s)
l	-litre
LB	-Luria broth
M	-Molar
mg	-milligram
min	-minute
ml	-millilitre
mol	-mole
ng	-nanogram
p	-plasmid
PCR	-Polymerase Chain Reaction
RNase	-ribonuclease
rpm	-revolutions per minute
sec	-second
SDS	-sodium dodecyl sulphate
T	-thymidine
TEMED	-N,N,N'-tetramethylethylenediamine
Tris	-Tris (hydroxymethyl) aminomethane
µg	-microgram
µl	-microlitre

UV	-ultraviolet
X-gal	-5-bromo-4-chloro-3-indol-β-D-galactopyranoside

Amino Acids

Amino acid	Three letter abbreviation	One letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Abstract

This work is concerned with the molecular analysis of ascomycete mating type genes of various *Sordaria* species. Work previously published has reported the cloning and characterisation of mating type genes from several *Neurospora* species. In heterothallic species the genotype at the mating type locus (mtA or mt**a**) determines the mating type. Homothallic species, which proceed through the sexual cycle without the need to mate, have no obvious mating types but molecular analysis has been used to demonstrate the presence of mating type genes in species with this life cycle.

Neurospora species and *Sordaria* species both belong to the Sordariaceae and are closely related. Several λ clones containing putative *Sordaria* mating type genes from heterothallic and homothallic species had been isolated previously using *N. crassa* mtA and mt**a** probes. In this study the mtA-1 gene of the heterothallic species *S. sclerogenia* was subcloned from a λ clone and sequenced. The equivalent gene from *S. equina* (a homothallic species containing only the mtA sequence) was also subcloned and sequenced. A λ clone for the species *S. fimicola* was found to hybridise with both the mtA and mt**a** probes. *S. fimicola* is a homothallic species containing mtA and mt**a** in the same nucleus. On sequencing the lambda clone it was found that the mtA and mt**a** genes are linked in this species.

All the *Sordaria* mtA-1 genes contained putative DNA binding domains, α domains. The mt**a**-1 gene sequenced from *S. fimicola* contained a putative HMG box. The *S. equina* mtA-1 gene was expressed in a sterile *N. crassa* mt**a** mutant and was found to restore mating type function to the mutant. The mtA-1 gene did not however confer homothallic behaviour on the recipient mutant.

S. equina and *S. sclerogenia* contain a 59bp common region following on from the mtA-1 gene which is conserved in both these species and in *Neurospora* species. A variable region continues on from the common region in *S. equina* and *S. sclerogenia* and in *Neurospora* species. The variable region can differ between species and between mating types.

The evolutionary relationships between *Sordaria* species and *Neurospora* species were examined also although the limited amount of data available means any conclusions reached are tentative.

Table of Contents

Chapter 1

Introduction

1 Introduction

This thesis is concerned with the analysis of mating type genes of various species of *Sordaria.* In this introductory chapter what is known about the mating types of other fungi will be reviewed. The budding yeast *Saccharomyces cerevisiae* is considered because this was the first mating type system to be described at the molecular level and because, although some of the details are not relevant to the organisation of mating type in *Sordaria* species, the mechanism of action of the gene products has been conserved throughout evolution. The yeast system provides a starting point towards understanding other systems. The mating type genes of the filamentous ascomycetes are described next. These are of direct relevance to the research carried out in this thesis. Finally the more complex basidiomycetes are considered, the mating type systems within this fungal group constitute the most complex mating type systems investigated to date.

1.1 The Budding Yeast *Saccharomyces cerevisiae*

How mating type genes control cell type and mating in *S.cerevisiae* has been researched extensively. This first section will briefly cover points, namely mating type proteins which can act as transcriptional activators/repressors, which are of relevance to control of the sexual cycle in the filamentous ascomycetes.

1.1.1 Proliferation and Transitions in the *S. cerevisiae* Cell Cycle

When plenty of nutrients are available, *S. cerevisiae* will proliferate in the mitotic cell cycle. A haploid cell duplicates its 17 chromosomes and distributes them between the mother cell and a daughter cell which is budded off. Two mating types, **a** and α, exist in *S. cerevisiae* and cells of opposite mating type readily fuse with each other. Proliferation is abandoned temporarily when cells of opposite mating type communicate via pheromones. **a** cells produce **a**-mating type factor, a pheromone which consists of 12 amino acids (Betz *et al,* 1987). α cells produce α–mating type factor, a pheromone consisting of 13 amino acids (Stotzler *et al* , 1976). The **a**-mating type factor informs α cells, which have a receptor for the **a**-mating type factor on their surface, of the **a** cells' presence and induces the α cell to prepare for mating. Conversely the α–mating type factor prepares **a** cells, which have a receptor for the α-mating type factor, for mating. The two cells fuse to produce an **a**/α diploid cell. Under appropriate enviromental conditions **a**/α cells can form an ascus from which the four spores, the haploid products of meiosis, are produced (Reviewed in Herskowitz, 1988; Rine, 1986 and Sprague *et al* 1983a).

1.1.2 Structure of the Mating Type Locus and Mating Type Switching

Whether a cell is **a** or α depends on which allele is resident at the mating type locus (MAT locus) on chromosome III. MAT**a** and MATα have been cloned (Hicks *et al*, 1979 and Nasmyth and Tatchell, 1980) and sequenced (Astell *et al* , 1981) and a 747 bp region called Yα was found to be unique to MATα and a 642 bp region called Y**a** was found to be unique to MAT**a**. Y**a** and Yα are surrounded by regions found in both mating types. From left to right, as conventionally written, the mating type cassette segments are the W (723 bp) and X (704 bp) regions, Y**a**/Yα, and the Z1 (239 bp) and Z2 (88bp) regions.

To the right and left of the MAT locus some distance away there are the HMR and HML loci respectively. HMR usually contains silent Y**a** information and HML usually has silent Yα information. The Y**a** and Yα elements at HMR and HML again are surrounded by regions homologous to both mating types, although HMR has no W or Z2 segments (Nasmyth and Tatchell, 1980). The purpose of these HML and HMR regions is to provide "storage loci" which enable an **a** cell to switch mating type to an α cell, or *vice-versa.* The α or **a** cassette at MAT is removed and replaced by information from HML or HMR *via* a transposition event. The composition of HML/HMR is not affected by this transposition event, the transfer of information is non-reciprocal. This "switching" ability will not be discussed in detail here as the filamentous ascomycetes which are investigated in this study do not have the ability to switch mating type (mating type interconversion is reviewed in Sprague *et al*, 1983a and in Hicks *et al,* 1979).

1.1.3 The α1-α2 Hypothesis

In a MAT**a** cell, **a** pheromone is produced and a receptor for the α pheromone is expressed. MATα cells produce α pheromone and a receptor for the **a** pheromone. These products enable the cells to mate and form a MATα/MAT**a** diploid cell in which receptors and pheromones are not produced so hence the diploid cell cannot mate. Unlike a haploid cell the diploid cell can undergo meiosis and produce spores. The specific characteristics of each cell were proposed by MacKay and Manney (1974) to be a result of factors encoded at the MAT**a** and MATα loci which control the expression of genes unlinked to the MAT locus necessary for mating.

MacKay and Manney (1974) found sterile (*ste*) mutations which were: a) linked to MATα (*ste*1); b) unlinked to the MAT locus but specific for MAT**a** cells (*ste*2); c) unlinked to the MAT locus but specific for MATα cells (*ste*3) and d) unlinked to the MAT locus but non-specific for mating type (*ste*4/5). Significantly, no class of *ste* mutants were found where the mutation is linked to MAT**a**. *Ste*2 and *ste*3 mutants were unable to respond to the pheromone produced by the opposite mating type, but the ability to produce pheromone was retained along with the ability to sporulate. One could see here how these mating type specific mutations are in genes that code for the receptors to pheromones from the opposite sex. Mutations linked to the MATα locus abolish the ability to respond to **a** pheromone and to produce α pheromone. This suggests a positive regulatory role for the MATα locus in expressing the α pheromone and **a** receptor genes.

Sterile **a** and α mutants that are incapable of sporulating were proposed to be the result of mutations at the MAT locus. One non-sporulating mutant derived from MAT**a** was able to mate at a low frequency to a wild-type MATα to produce a non-sporulating diploid, unusually this could mate as an α strain. This diploid was mated with a diploid homozygous for **a** to produce a tetraploid with two normal **a** alleles, a normal α allele and the MAT allele from the original mutant. This tetraploid was able to sporulate. MacKay and Manney (1974) pointed out that if the original mutation was not at MAT and had its effect through a cytoplasmic product to produce the non-sporulating phenotype in the diploid, then the same phenotype should be apparent in the tetraploid. If this mutation was not at MAT and had a recessive action then the non-sporulating phenotype would not be present in both the diploid and the tetraploid. The non-sporulating phenotype here was proposed to be a result of a mutation carried at MAT**a**. Tetraploids made by mating a diploid formed from a non-sporulating mutant derived from MATα and a normal MAT**a** to a diploid homozygous for MATα were also found to sporulate.

MacKay and Manney (1974) proposed that the mating type alleles are regulatory in function and regulate most if not all of the genes necessary for mating. Genes under the control of MAT were split into four groups, those expressed only in MAT**a** cells, those expressed only in MATα cells, genes expressed in both MATα and MAT**a** cells and genes expressed only in heterozygous diploids.

Four mutations of the MATα locus isolated by MacKay and Manney (1974) were investigated in greater detail by Strathern *et al* (1981). Three of the mutations led to sterility but in those rare diploids formed, sporulation could take place. The fourth mutation also led to sterility but rare diploids were unable to sporulate. The two types of mutation at MATα were found to define two complementation groups, α1 and α2. α1 mutants are able to sporulate when part of a diploid, do not secrete α-pheromone or respond to **a**-pheromone. α2 mutants are unable to sporulate when part of a diploid, do not produce α-pheromone or respond to **a**-pheromone but in contrast to α1 mutants, they display some MAT**a** cell characteristics. α2 mutants respond to α-pheromone and produce **a**-pheromone.

Mutations at MAT**a** appear to have no effect on the mating activity of MAT**a** haploids but do cause a failure of sporulation when a cell mutant at MAT**a** forms a diploid with the opposite mating type. This results in the diploid mating as an α cell. Mutations at MAT**a** were classed as **a**1 mutations. From the observed phenotypes of MATα and MAT**a**

mutations a model, the α1-α2 hypothesis, was proposed (Strathern *et al*, 1981) for the control of cell type by the MAT loci.

1. MATα1 is a positive regulator of α-specific genes like α-pheromone and α-specific STE genes. Mutations of MATα1 cause sterility because MATα1 specific genes are no longer expressed.

2. MATα2 encodes a negative repressor of **a**-specific genes. Mutations at MATα2 cause the MATα cell to have a phenotype similar in some ways to a MAT**a** cell but α-specific functions will still be expressed due to the presence of a functional MATα1. The antagonism between the α and **a**-specific functions is probably the reason why MATα2 mutants are deficient in mating. If a MATα/MAT**a** diploid carries a mutation at MATα2 then the diploid will be deficient in sporulation functions. MATα2 plays a regulatory role necessary for sporulation to take place.

3. MAT**a**1 is also necessary for sporulation function in diploids. This makes MAT**a** and MATα co-dominant alleles as both **a**1 and α2 must be present for sporulation in diploids. Together MAT**a**1 and MATα2 also block mating in diploids by inhibiting MATα1 expression. MAT**a** cells have a MAT**a** phenotype due to the absence of MATα2.

The α1-α2 hypothesis is summarised in figure 1.1.3.

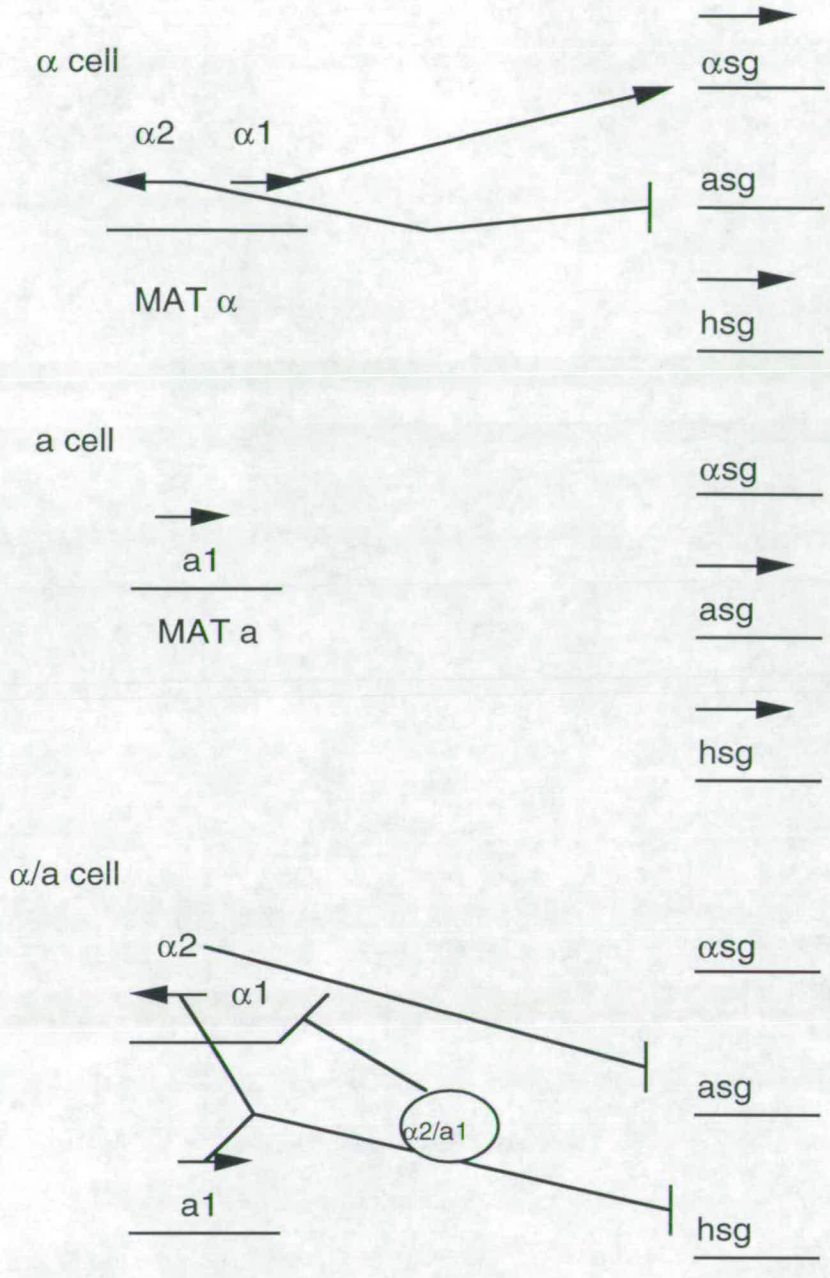


FIGURE 1.1.3

MAT genes as master control loci in α haploid , a haploid and α/a diploid cells. The effect of the proteins on α-specific genes, **a**-specific genes and haploid-specific genes is shown. Where a protein has a positive effect on a target gene is denoted by an arrow. Vertical bars denote blocked processes. α2/**a**1 blocks the expression of haploid-specific genes allowing sporulation to take place. αsg=α-specific genes, **a**sg=**a**-specific genes, hsg=haploid specific genes.

1.1.4 The Mutations at the MAT Alleles can be Mapped to Transcripts Expressed at the MAT Alleles.

The α1-α2 hypothesis proposes that MATα and MAT**a** encode proteins that regulate the transcription of genes that play a role in mating and sporulation. At MATα there are two mRNAs transcribed divergently. One, α1, is transcribed from the α–specific sequence Yα and the other, α2, is encoded at the X region common to both mating types. At MAT**a** an **a**1 mRNA is encoded at the **a**-specific DNA sequence, Y**a**, but a second mRNA, **a**2, is transcribed from the X region common to both MAT**a** and MATα (Astell *et al* , 1981). Tatchell *et al* (1981) have connected the MAT loci transcripts with the MATα1, MATα2 and MAT**a**1 complementation groups. Using *in vitro* mutagenesis new XhoI restriction sites were introduced into cloned DNA containing either the MAT**a** or the MATα locus. This technique resulted in deletions and duplications being produced at the site of the new restriction site.

Twelve MAT**a** plasmids containing Xho linker mutations were tested for their ability to complement a MAT**a**1 mutant strain which had a failure in sporulation when part of a diploid and allowed the diploid to mate as an α strain. Each of the twelve plasmids were used to transform a mutant MAT**a**1 diploid, mat**a**1/MATα. Four of the mutant plasmids with mutations in the **a**1 gene failed to restore sporulation functions in the diploid and failed to stop the diploid mating as an α strain, so it was concluded that the **a**1 protein is involved in activating sporulation and repressing α mating in diploids. To assign a possible function to the **a**2 transcript, the twelve plasmids with mutations within MAT**a** were transformed into MATα/MATα diploids to see if the plasmids could complement the sporulation and mating repression defects. Unlike those plasmids with mutations in **a**1 the plasmids with mutations within **a**2 were able to complement the defects, even those plasmids with large deletions at **a**2. No function could be assigned to **a**2 by this complementation assay.

Thirty-nine MATα mutant plasmids were tested for their ability to complement MATα1 and MATα2 mutations. Thirty-three of the mutants complemented MATα1 mutations, the remaining six mutant plasmids, which failed to complement, all had mutations in the α1 coding region. Eight mutant plasmids failed to complement the sporulation defect in MAT**a**/matα2 diploids and failed to complement the mating defect in matα2 haploids. All eight plasmids had mutations in the α2 coding region. The α1 protein is responsible for expression of α-specific genes necessary for mating, while the α2 protein represses **a**-

specific genes in haploids and along with the **a**1 protein, activates sporulation in diploids.

1.1.5 The MAT Proteins Regulate the Transcription of Cell Type Specific Genes

The STE3 gene was cloned when a yeast genomic clone bank was screened and a plasmid found that contained the information to complement the mating defect of *ste* 3 mutants (Sprague *et al*, 1983b). The STE3 gene is thought to encode the α-specific receptor for the **a**-factor (Hagen *et al*, 1984). A matα2/matα1 mutant cell would mate as an **a**-cell as the mutant α1 cannot express α-specific genes and the mutant α2 cannot repress **a**-specific genes. The same phenotype was shown in a matα2/MATα1 *ste*3 cell. This suggested that STE3 is positively regulated by α1 and that STE3 is required for α-specific mating activity. Sprague *et al* (1983b) used the STE3 gene as a probe to find out if mutant matα1 cells produce STE3 RNA. The STE3 probe showed the presence of STE3 RNA in MATα1 cells but not in mutant matα1 cells. No STE3 RNA was detected in **a** or **a**/α cells. STE3 is expressed only in α-cells and requires the MATα1 product for expression.

The STE2 gene was cloned by complementation of the *ste*2 mating defect. Using the STE2 gene as a probe, STE2 RNA was only detected in **a** or matα2 cells (Hartig *et al,* 1986). The MATα2 gene product represses expression of STE2 so no STE2 RNA was found in α or **a**/α cells. STE2 RNA levels were increased in **a** cells that had been exposed to α-factor. STE2 has been proposed to encode the **a**-specific receptor for the α–factor (Jenness *et al,* 1983).

The STE12 gene was cloned by complementation of the *ste*12 mating defect. The STE12 protein is required by both **a** and α cells to mate. The STE12 recognition sequence on DNA is called a pheromone response element (PRE) and STE12 upregulates **a** and α-specific genes in response to pheromone stimulation (Dolan and Fields, 1991). STE12 transcripts were present in **a**, α and **a**/α cells but at a reduced level in **a**/α cells. **a**/α strains mutant at α2 or **a**1 were probed for STE12 RNA and the repression of STE12 expression was found to be abolished in these strains. In the diploid state the combined activity of the MAT**a**1 and MATα2 products represses STE12 expression (Fields and Herskowitz, 1987).

1.1.6 The MAT Proteins Regulate the Transcription of Cell-Type Specific Genes as Part of Protein Complexes Formed with Other Regulatory Proteins

1.1.6.1 The Yeast Transcriptional Activator MCM1 is Involved in Both a and α-Specific Gene Expression.

Jarvis *et al* (1987) found that the control region of the α-specific gene STE3 contains a 26bp sequence found at the control regions of other α-specific genes. The 26bp sequence has a P element, 16bp and an imperfect palindrome, and a Q element, a 10bp sequence thought to be the binding site for α1 (Bender and Sprague, 1987). The α1 protein was shown by Bender and Sprague (1987) to bind to the QP element only in conjunction with MCM1. MCM1 and α1 bind cooperatively to the QP element, the complex formed as a dimer of MCM1 and a single molecule of α1 (Primig *et al* , 1991). MCM1 on its own binds only weakly to the imperfect palindromic P elements found at α-specific genes, α1 allows tighter binding. P elements at **a**-specific genes were found to be more symmetrical, like the palindromic 14bp version of the P element [P(PAL)] synthesized by Jarvis *et al* (1987). [P(PAL)] was used to replace the upstream activation sequence (UAS) of the CYC1-lacZ gene to see if the gene could be expressed in **a** and α cells, which it was. In contrast the P element of STE3 had no UAS activity in **a** and α cells. MCM1 interacts with P(STE3) in conjunction with α1 but can recognize P elements that are perfectly palindromic, like those found at **a**-specific genes. Elble and Tye (1991) have shown that a mutant allele of MCM1 affected both the activation and repression of **a**-specific genes. It was originally thought that the α2 protein caused repression of a-specific genes by blocking the MCM1 binding sites (Bender and Sprague, 1987). At **a**-specific genes the MCM1 binding sites is flanked by α2 binding sites and Keleher *et al* (1988) showed that MCM1 and α2 bind cooperatively to the α2 operator.

1.1.6.2 STE12 is Involved Also in a and α-Specific Gene Expression

Errede and Ammerer (1989) found, using band-shift assays, that the STE12 protein along with MCM1 bound to the STE2 UAS. In DNA protection assays STE12 left a footprint at the PRE element present at the STE2 UAS. Through weak protein-protein contact STE12 and MCM1 bind cooperatively to the DNA (Bruhn and Sprague, 1994). At α-specific genes no good matches for the PRE element were found so it was difficult

to explain how activation by STE12 occurred. Yuan *et al* (1993) showed that bacterially expressed STE12 and α1 could associate *in vitro* so STE12 could act through a protein-protein interaction with α1.

1.1.6.3 SSN6 and TUP1 are Part of the Complexes that Repress a-Specific Genes and Haploid Specific Genes

Haploid specific genes like STE12 are repressed in the diploid state by the cooperative binding of **a**1 and α2 proteins to the haploid-specific gene operator (Goutte and Johnson, 1993). Keleher *et al* (1992) showed that diploid cells homozygous for either SSN6 or TUP1 mutations sporulated poorly due to haploid specific genes not being repressed. Disruption of the SSN6 or TUP1 genes in α cells lead to the inappropriate expression of **a**-specific products. SSN6 and TUP1 are important in the repression of **a**-specific and haploid-specific genes.

The regulation of α, **a** and haploid specific genes as currently understood is described in figure 1.1.6 taken from Johnson (1995).

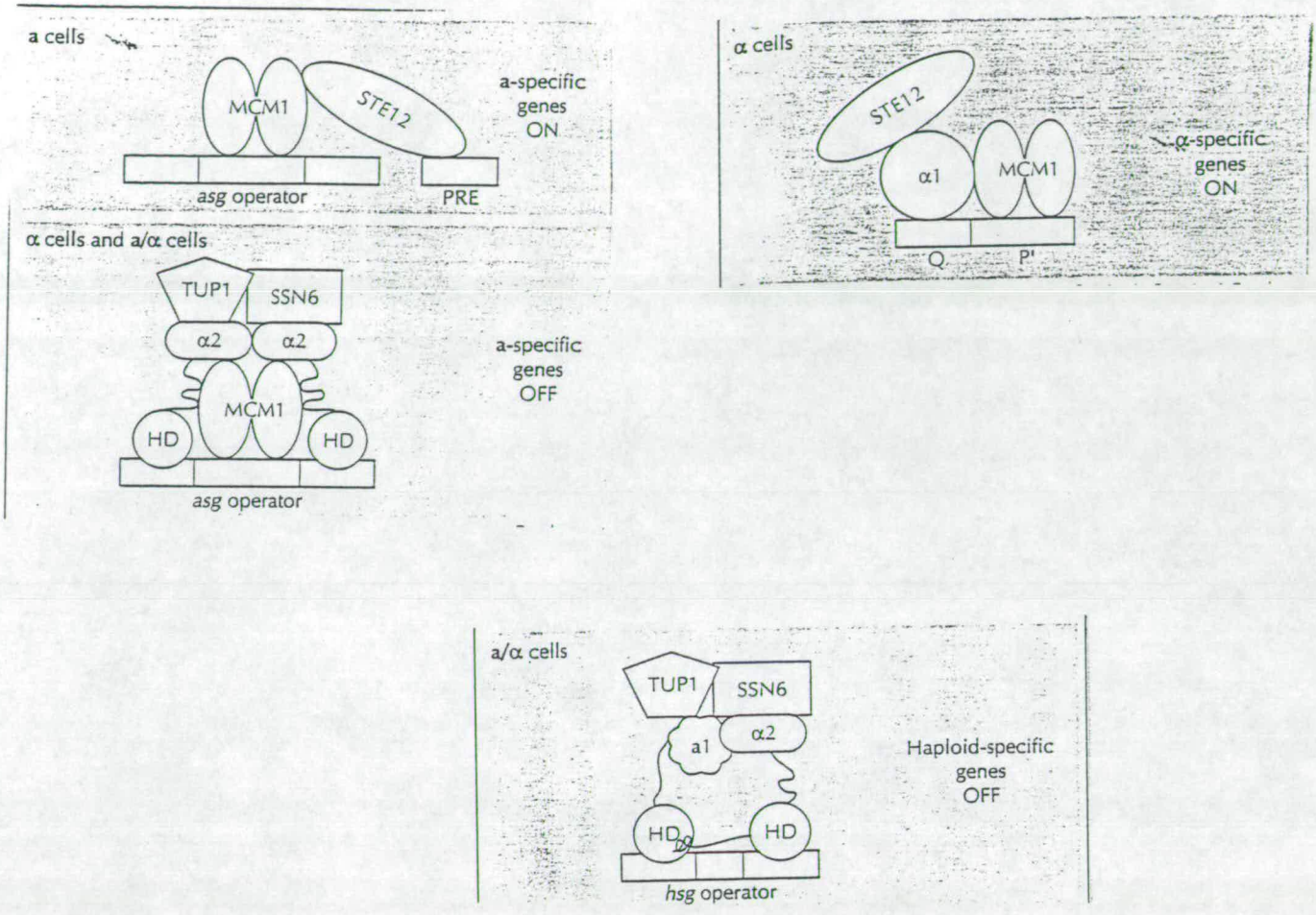


FIGURE 1.1.6

Cell type regulation of **a**-specific genes, α specific genes and haploid specific genes.

1.2 The Filamentous Ascomycetes

The two species of Ascomycetes which will be discussed in detail in this section are *Neurospora crassa* and *Podospora anserina*. The life cycle of *N. crassa* will be described as an example of a fungal Ascomycete life cycle (Fincham, 1983).

1.2.1 Life Cycle of *N. crassa*

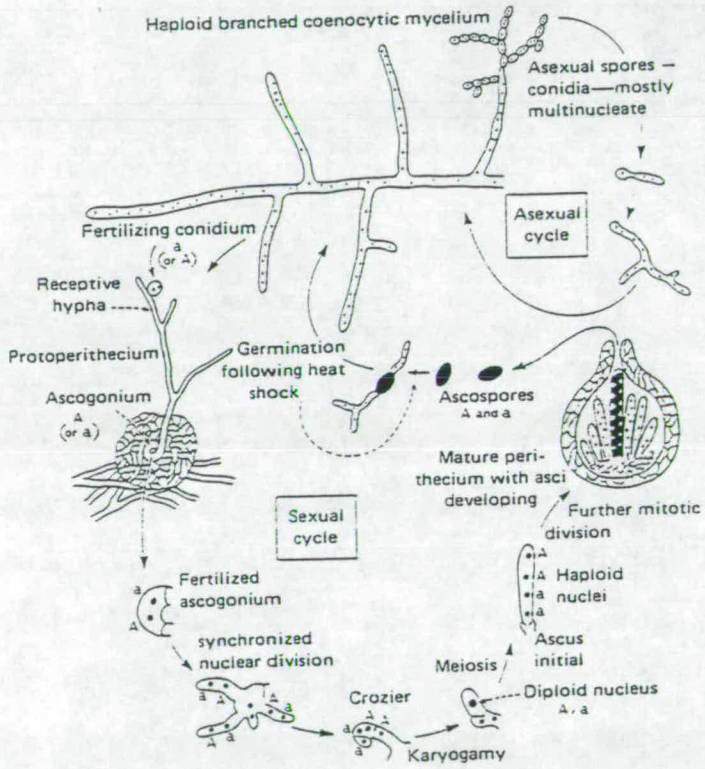


FIGURE 1.2.1

The life cycle of the filamentous Ascomycete *N. crassa*. The protoperithecium and perithecium are shown as if they are sectioned vertically.

N. crassa is a heterothallic species, that is individuals are either A mating type or **a** mating type with A cells mating to **a** cells (Shear and Dodge, 1927). *N. crassa* forms branching filaments or hyphae in its vegetative state during the asexual cycle. The hyphae are made up of multinucleate cells. If the hyphae of opposite mating type fuse during the asexual cycle then a process called heterokaryon incompatibility occurs which results in protoplasmic incompatibility (Beadle and Coonradt, 1944). This process was described in detail by Garnjobst and Wilson (1956). Hyphae differing in mating type were fused on slide preparations. Protoplasmic exchange in the vicinity of the fusion was followed by death of the the fused cells. Garnjobst and Wilson (1956) suggested that this protoplasmic incompatibility between mtA and mt**a** individuals of different species may have contributed to speciation in *Neurospora.* The mating type factors A and **a** are incompatibility factors and it was noted by Garnjobst and Wilson (1956) that protoplasmic incompatibility only occurred during the vegetative stages.

For the sexual cycle to be initiated, A and **a** cultures must come together under suitable conditions. Conditions of nitrogen starvation and a low temperature of 25° induce the formation of the female reproductive structure, the protoperithecium. Both mating types are hermaphroditic and can form male and female structures. The protoperithecium has a specialised hypha, the trichogyne, which grows towards the male cell. The trichogyne is attracted to the male cell of opposite mating type by a diffusible pheromone-like substance secreted from the male cell. (Bistis, 1981; Bistis, 1983) The protoperithecium also contains a coiled cell called the ascogonium. The nucleus from the male conidium passes via the trichogyne into the ascogonium where it associates with a nucleus of opposite mating type. After a period of synchronized nuclear division the paired nuclei fuse (karyogamy) to give diploid cells. This is immediately followed by meiosis to give four haploid nuclei which undergo a mitotic division to give eight nuclei. Eight ascospores contain the haploid products of meiosis. Mature ascospores are induced to germinate after heat shock which returns the fungus to the vegetative state (reviewed in Fincham, 1983; Metzenberg and Glass, 1990).

1.2.2 Structure and Functions of the Mating Type Idiomorphs

1.2.2.1 *Neurospora crassa*

Whether a heterothallic organism is A or **a** is determined by the allele carried at the mating type locus. The mating type genes were cloned by Glass *et al* (1988). MtA DNA was cloned on the basis that *un-3*, a temperature-sensitive mutation, is the closest selectable marker to mtA. Tansformation with cosmid pSV6:10A from a genomic mtA library (Vollmer and Yanofsky, 1986) allowed the *un-3* recipient to grow at the restrictive temperature. Sterile cultures of *N. crassa* , carrying a mutation at the mating type locus, were transformed with pSV6:10A and the recipient fungi were able to mate as mtA cells showing that a functional mtA gene was present on pSV6:10A. A 4.4kb fragment from pSV6:10A was found to be unique to mtA cells. Transformation experiments showed that the fragment contained all the information needed to confer mtA activity on a mating type mutant recipient. A λ library made from randomly sheared fragments from a mt**a** strain was probed with pSV6:10A and sequences that were common to mtA and mt**a** and flanked the DNA unique to mtA were detected. DNA that bound to the probe was found to confer mt**a** mating type behaviour to a mutant mtA recipient. Mt**a** DNA unique to mt**a** cells and responsible for conferring mt**a** mating type behaviour was found to reside on a 4.5kb fragment. Not all of the mt**a** and mtA specific DNA is required for mating. A 1.7kb *Pst* I-*Sal* I fragment from the 4.4kb mtA-specific fragment was found to be sufficient to confer mtA mating type function on a sterile recipient. Similarly a 2kb *Eco* RV-*Bam* HI fragment from the 4.5kb mt**a**-specific fragment was able to confer mt**a** mating type function on a sterile recipient. Probes were made from the A-specific and **a**-specific mating type regions which would not hybridise to genomic DNA from the opposite mating type. Thus the genome of each mating type does not contain silent mating type genes as found in *S. cerevisiae* (Glass *et al* 1988). It was suggested that the two mating type genes be termed "idiomorphs" as one is not derived from the other by simple mutation as is the case with alleles.

Glass *et al* (1990a) and Staben and Yanofsky (1990) have sequenced the mtA and mt**a** mating type regions respectively. The mtA idiomorph was found to be 5301bp long and the mt**a** idiomorph 3235bp long. A 928bp long open reading frame, mtA-1, was identified in the A idiomorph. The **a** idiomorph has an ORF, mt**a**-1,which is 1260bp long. The polypeptides encoded by these ORF's were shown by transformation experiments to contain the mating and vegetative incompatibility functions. Ectopic integration of fragments containing the mtA-1 and mt**a**-1 genes conferred on the recipient

the ability to mate as a mtA or mt**a** strain, but if the transformants were crossed no ascospores are produced. Vegetative incompatibilty functions can be assayed by transforming mt**a**-1 into mtA cells and *vice-versa*. When a transforming construct contains the ability to cause vegetative incompatibility then the transformation efficiency is at least 20-fold lower than when transforming mtA-1 into mtA cells and mt**a**-1 into mt**a** cells (Glass *et al* , 1988). Figure 1.2.2.1 a) shows the relationship between the mt**a** idiomorph and the mtA idiomorph with respect to their highly conserved flanking DNA (Glass *et al*, 1990a).

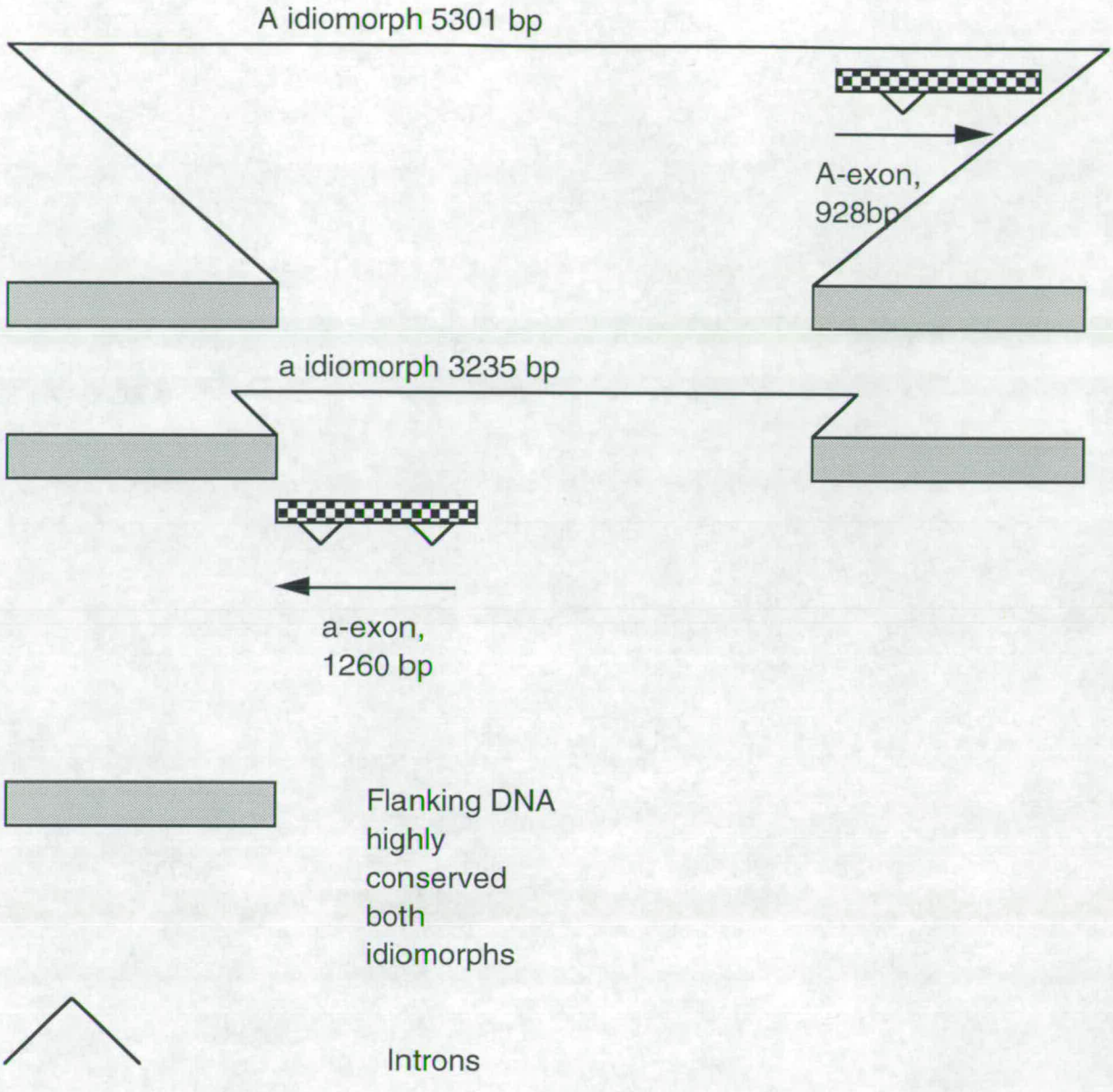


FIGURE 1.2.2.1a)

Comparison between the mtA idiomorph and the mt**a** idiomorph. The transition from the dissimilar idiomorphs to conserved flank is abrupt. The checked boxes indicate the mtA-1 and mt**a**-1 open reading frames. Their orientation is shown by arrows.

The mt A-1 polypeptide consists of 288 amino acids and amino acids 45-59 show similarity to amino acids 90-104 of the MAT α1 polypeptide. This region has been called the α-domain, a DNA binding domain (Glass *et al*, 1990a). This suggests that mtA-1 may be a transcription factor. The amino terminal half of the mt**a**-1 polypeptide shows similarities to the shorter *Schizosaccharomyces pombe* Mat-Mc polypeptide. This region of *S.pombe* contains an HMG box motif found in the human nucleolar transcription factor hUBF (Jantzen *et al* , 1990).

The functional regions of mtA-1, the incompatibility and mating functions, have been separated by Saupe *et al* (1996). A new mutant, A^{m99} , was obtained which had lost the vegetative incompatibility function but had retained the ability to mate as the protoperithecial "female" parent. The DNA sequence of the mutant was determined and it was found that the mtA-1 of the mutant was truncated after the first 85 amino acids, suggesting that the N-terminal region is sufficient for female fertility functions. MtA-1 deletion constructs were also made and assayed for fertility functions via transformation experiments. Transformants containing constructs which retained amino acids 1-227 were able to mate as mtA cells. Male mating function was lost for a construct which was truncated at amino acid 184, indicating a requirement for amino acids 184-227 for full mating activity. A^{m} mutant alleles that had been previously characterised were cloned by PCR and transformed onto mt**a** cells so it could be determined if the mutant gene elicits an incompatibility reaction. A^{m64} and A^{m54} constructs were found to give lower transformation efficiencies with a mt**a** strain, which was interpreted as originating from the vegetative incompatibility function. A^{m54} has a frameshift mutation after the first 163 amino acids and A^{m64} after the first 111 indicating that the first 111 amino acids contain the information to produce an incompatibility reaction when mtA-1 and mt**a**-1 reside in the same nucleus. A^{m42}, which has a frameshift mutation after the first 100 amino acids, does not elicit an incompatibility reaction when transformed into a mt**a** strain which suggests a region between amino acid 100 and 111 is required for the incompatibility process. Specific amino acids could be mutated to confirm this region is required for the vegetative incompatibility but these experiments have yet to be reported.

Philley and Staben (1994) conducted *in vivo* and *in vitro* functional studies on mt**a**-1 and mutant derivatives. The mt**a**-1 polypeptide was expressed in *E-coli* and tested for binding to the A and a idiomorphs by gel mobility shift assays. Surprisingly the polypeptide was found to bind to fragments within and surrounding both idiomorphs. The mt**a**-1 polypeptide was found to protect a CTTTG element from DNaseI digestion, an element found in the binding sites of other HMG box polypeptides. Seven CTTTG

sequences cluster between nucleotides 2750 and 3250 of mtA, the region that bound mt**a**-1 in mobility shift assays. This region of mtA corresponds to the transcript mtA-2 (see below). Perhaps mt**a**-1 regulates the expession of mtA-2. Mt**a**-1 mutations were tested for their binding properties *in vitro* and it was found that all mutants that retained an intact HMG box retained the ability to bind DNA. An $\mathbf{a}^{m30}$ mutant, where the HMG box contains a large insertion knocking out mating ability, was transformed with wild-type mt**a**-1 and also with mutant derivatives of mt**a**-1 and scored for ability to mate with mtA. It was found that those transformants with both an intact HMG box and a portion of the carboxyl terminal region were able to mate. It is proposed that, as the carboxyl terminal region is acidic and proline rich and that mutant polypeptides lacking acidic tails do not confer mating activity, the acidic tails must have a function required for mating. Mutations within the HMG box did not interfere with vegetative incompatibility *in vivo*. Deleting amino acids 216-220, downstream of the HMG box, eliminated vegetative incompatibility but did not interfere with mating or DNA binding. The vegetative incompatibility process therefore functions via a mechanism separate to that of mating and DNA binding.

MtA-1 resides in a 1kb portion of the 5.3kb mtA idiomorph. It was possible the idiomorph contained other regions with functional importance, given the size of the idiomorph. Mt**a**-1 is the only ORF encoded in the mt**a** idiomorph and contains all the necessary information for completion of the sexual cycle. Glass and Lee (1992) used repeat induced point (RIP) mutation to identify other regions in mtA which are required in the sexual cycle. RIP works in *N. crassa* by mutating duplicated stretches of DNA in the haploid nuclei of the heterokaryotic tissue formed after fertilisation (Selker, 1990). The mutations are in the form of G-C to A-T transitions and often sequences altered by RIP are methylated.

Since most transformants in *N. crassa* are ectopic i.e they do not replace the resident DNA sequence but instead form an additional copy, RIP forms an experimentally convenient method for mutating any given stretch of DNA and thus analysing the consequences of the mutation. Glass and Lee (1992) took a fragment of the mtA idiomorph 3.0kb in length containing sequences upstream of mtA-1. This fragment was transformed into mtA spheroplasts so the 3.0kb fragment would be duplicated. The transformants were then crossed to a mt**a** strain to produce mtA RIP mutants. These mtA mutants were capable of mating but very few ascospores were produced as a result of crossing the mutant with mt**a**. Ferreira *et al* (1996) identified two genes, mtA-2 and mtA-3, upstream of mtA-1. The 3.0kb DNA stretch affected by RIP in the mutant was

sequenced and it was confirmed that the affected area encompassed all of mtA-2 and 700bp of mtA-3. Fifty-two G-C to A-T transition mutations were identified in mtA-2 but only six RIP mutations were found in mtA-3. It is likely that the ascospore deficient phenotype of the RIP mutant is due to the mutations in mtA-2.

The relative positions of mtA-2 and mtA-3 to mt A-1 are shown in figure 1.2.2.1b) (Ferreira *et al*, 1996). The 70bp between the two genes contains repeated sequences which may indicate that the two genes are coordinately controlled by specific factors. The mt A-2 peptide is 373 amino acids long and the mt A-3 peptide is 324 amino acids long. Mt A-3 has an HMG domain. It is likely that mt A-2 and mt A-3 are transcription factors controlling post-fertilisation events necessary for ascospore production. Northern and RT-PCR analyses have revealed however that mt A-2 and mt A-3 are constitutively expressed during the vegetative and sexual cycles. Upstream of the mt A-2 and mt A-3 transcripts there are small upstream ORFs (uORFs) which may affect translation. So perhaps there is mechanism which allows mt A-2 and mt A-3 to be selectively translated during the sexual cycle. Other regulatory uORFs have been described in *N. crassa,* in *Cpc-1*, the homolog of *GCN4* (Paluh *et al,* 1988) and in *arg-2* (Orbach *et al,* 1990). Alternatively perhaps the products of mtA-2 and mtA-3 interact with other proteins in the perithecia to regulate post-fertilization events.

N.crassa

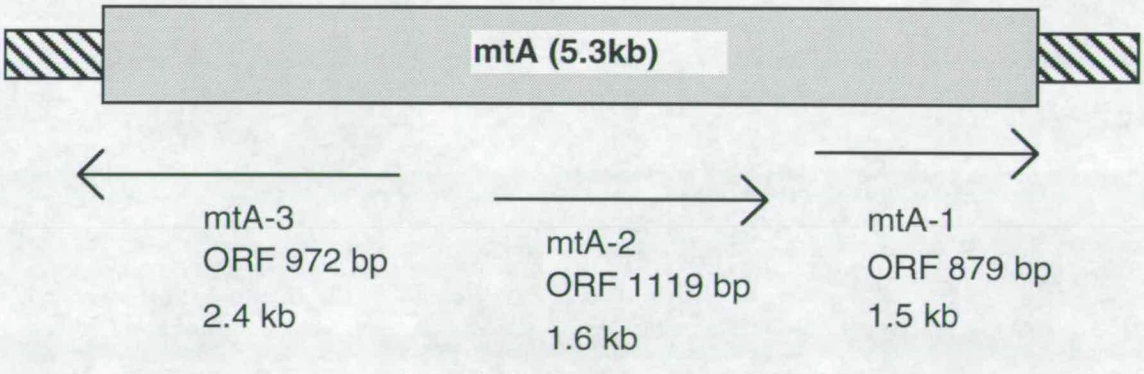


FIGURE 1.2.2.1b)

The three genes of the mtA idiomorph of *N. crassa*. The orientation of the three transcripts is indicated by the arrows. Flanking DNA is shown as hatched boxes.

Nelson and Metzenberg (1992) have identified fourteen genes involved in sexual development (*sexual development [sdv]* genes)), most of which require a functional mt A-1 mating product for expression. To isolate these genes strains of *N. crassa* were grown under conditions (nitrogen starvation, 25°, bright light) which encourage the development of the female reproductive structure. Strains were harvested after 2-4 days so that transcripts specific to sexual development would be present. Polyadenylated RNA was isolated and labelled cDNAs were synthesised using random primers and M-MLV reverse transcriptase. The labelled cDNAs were hybridised with an excess of mRNA isolated from a vegetatively growing culture so that any single-stranded cDNAs would not correspond to transcripts present during vegetative growth. These single-stranded cDNAs were used to probe a cosmid library containing large inserts of genomic *Neurospora* DNA. Fifty strongly hybridising cosmids were selected for further study. These cosmids were used to probe mRNA from a vegetatively growing strain and mRNA from a strain grown under crossing conditions. Thirty five of the cosmids were shown to encode transcripts whose levels were increased and sometimes only present under crossing conditions. These 35 cosmids were found in turn to contain about 30 sexual development genes. Many of the genes appear to be clustered as some of the cosmids contain two or three genes that are expressed only in conditions favouring sexual development. Fourteen of the *sdv* genes were subcloned, and their expression in a mutant A^{m44} strain, grown under crossing conditions, was examined. The A^{m44} strain has a frameshift mutation in mtA-1 which inactivates the mating type function of the protein (Glass *et al*, 1990a). The *sdv* genes were not expressed in significant amounts in A^{m44} which indicates that expression of these genes is dependent on a functional mtA-1 product. It is not known whether mtA-1 acts on the *sdv* genes directly or acts through intermediate genes. The effect of mt**a**-1 on the expression of the *sdv* genes is under investigation (Nelson and Metzenberg, 1992).

1.2.2.2 *Podospora anserina*

P. anserina is closely related to *N. crassa.* The sexual cycle of *P. anserina* is almost identical to that of *N. crassa.* except nuclei of opposite mating type are compartmentalised within a single ascospore so that a single ascospore gives rise to a self fertile culture following germination. The mating type locus of *P. anserina* therefore does not show vegetative incompatibility function.

The two mating types genes of *P. anserina* are called mat+ and mat- (Fincham *et al*, 1979) . Mat- was cloned using the mtA mating type probe from *N. crassa* under non-stringent conditions and mat+ cloned on the basis that the mat- flanking sequence will bind that of mat+ as they share highly homologous flanking DNA (Picard *et al*, 1991). Mat+ and Mat- regions were found to be 3.8kb and 4.7kb respectively. Like A and **a** of *N. crassa,* mat- and mat+ are unlike in sequence and so can be termed idiomorphs. Single ORF's are encoded at mat+ and mat- that can restore mating type function when transformed into sterile mutants. The mat- ORF was named FMR1, for fertilisation minus regulator and was proposed to encode a polypeptide of around 305 or 349 amino acids depending on where the intron splice sites are located. As the *N. crassa* mtA-1 gene was used to clone mat-, similarities between the two genes were expected. FMR1 and mtA-1 share 106 identical amino acids out of 196 amino acids at the N-terminus but the polypeptides are completely dissimilar at the C-terminal end. FMR1 like mtA-1 also shows similarities to the yeast α-1 protein indicating that FMR1 could be a transcription factor. The shared region of homology for these three proteins was termed the α-domain and was proposed to control genes involved in mating (Debuchy and Coppin, 1992; Glass and Kuldau, 1992).

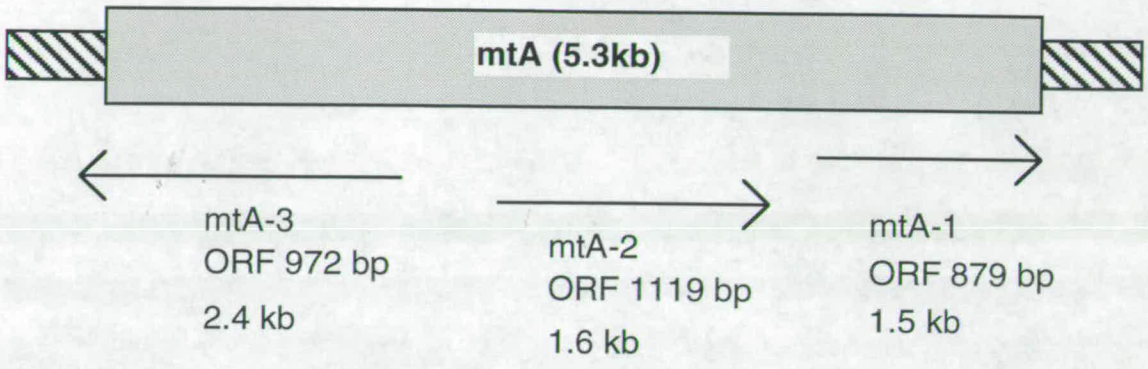
The mat+ ORF was called FPR1 for fertilisation plus regulator. The 365 amino acid FPR1 protein shows homology to *N. crassa* mt**a**-1 over a stretch of DNA encoding an HMG domain. This suggests FPR1 is also a transcription factor.

Transformation experiments using subclones derived from the mating type regions of a mat- strain showed that FMR1 determined mating type identity but that additional sequences were required for the development of mature fruiting bodies. Debuchy and Coppin (1992) suggested that there were at least two other genes, encoded at mat-, controlling post fertilization events. FPR1, like mt**a**-1, seems to encode all the information necessary for post fertilization events; there seems to be no additional information found in the mat+ idiomorph. Debuchy *et al* (1993) have characterised two

more genes encoded at mat- along with FMR1. These genes were called SMR1 (Sporulation Minus Regulator 1) and SMR2 (Sporulation Minus Regulator 2). Experiments showed that FMR1 is also required for post fertilization events. A deletion covering the 3' end of FMR1 and the downstream portion of mat- did not cause infertility but impaired ascospore formation. The defect in ascospore formation could have been the result of the deletion at the 3' end of FMR1 or a deletion in a gene downstream of FMR1. The defect was localised to the 3' end of FMR1 by cotransforming a sterile mutant, deleted for the mat locus, with a plasmid carrying SMR2, SMR1 and FMR1 deleted at the 3' end along with a plasmid carrying the complete FMR1. The cotransformation resulted in the mutant behaving like a wild type strain, the complete FMR1 was enough to correct the defect in ascospore formation. So it is FMR1, not a gene adjacent to its 3' end that is involved in post-fertilization events.

Mutational analysis demonstrated that SMR1 and SMR2 were necessary for post fertilisation events. The gene product of SMR1 has an acidic domain found in other transcription factors. An HMG domain was found to be present in the deduced SMR2 polypeptide. SMR1 showed 23% identity to mtA-2 and SMR2 showed 22% identity to mtA-3 at the amino acid level (Ferreira *et al* , 1996). The positions of the genes of mat - compared with the positions of the genes of mtA in *N. crassa* are shown in figure 1.2.2.2 (Ferreira *et al* , 1996) .

N.crassa



P. anserina

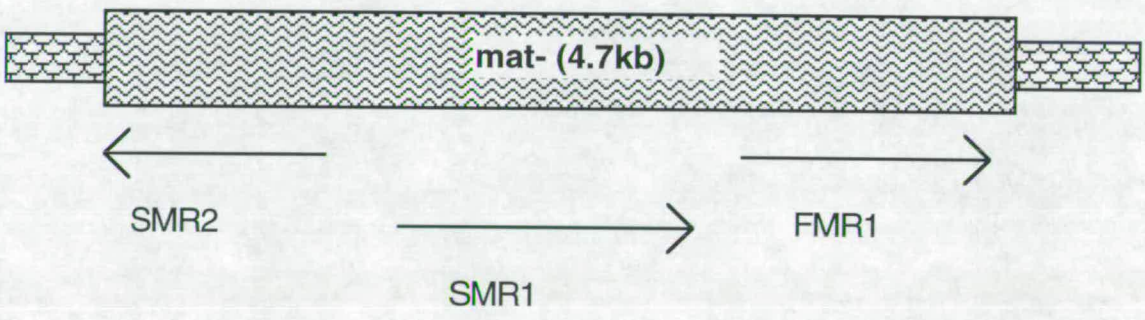


FIGURE 1.2.2.2

The positions of the genes in the *N. crassa* mtA idiomorph compared with the position of the genes in the *P.anserina* mat- idiomorph. Arrows show the orientation of the transcripts. Hatched and scalloped boxes indicate flanking DNA.

Zickler *et al* (1995) investigated the post fertilization role of the mat+ and mat- genes. Mutant versions of the four mat genes of mat- and mat+ were transformed into a strain deleted for its mat locus so that some transformants carried the mat+ mutated gene and the other transformants carried all three mat- genes, but with only one of the three genes mutated. The mat- FMR1-1, SMR1-1, SMR1-2 and SMR2-1 mutations were frameshift mutations. The FPR1-1 mutation corresponds to a deletion at the 3' end of the gene. FPR1-1 and FMR1-1 still retained their fertility functions. Each mat mutant transformant was crossed to a strain carrying the wild-type compatible mating type and the crosses observed for their effect on post fertilisation events. In mutant X wild type crosses, some croziers contained only one haploid nucleus. Croziers are always binucleate in wild type crosses (see Figure 1.2.1). Where a mutant X wild type cross did result in a binucleate crozier, the crozier sometimes contained two nuclei from the mutant parent. Uninucleate croziers result in a haploid meiosis and abnormal spore formation. So during post fertilisation events an important step is the cellularisation of two nuclei, one from each parent, in the crozier. Mutations in the mating type genes can result in selfish behaviour of the mutant nucleus so it ignores its wild type partner. The mat genes ensure the resulting spores are dikaryotic and biparental. When the mat- mutant strains were crossed to a wild type strain, SMR1 mutants showed the strongest effects on events. Uniparental progeny were always produced. SMR2 and FMR1 mutants still produced biparental progeny although at a lower frequency than that found in wild type crosses. The cytoskeleton has been proposed to play a part in the control of nuclear isolation (Thompson-Coffe and Zickler, 1994). Zickler *et al* (1995) suggested that the wild type nucleus does not migrate to the crozier cortex as it cannot establish a microtubule bridge with the mutant nucleus. The mutant nucleus however can migrate alone to the cortex without the aid of the bridge.

Regardless of the precise mechanics of the action, it is evident that the mtA-1 mtA-2 and mtA-3 genes in *N. crassa* and FMR1, SMR1 and SMR2 in *P. anserina* play an important role in both fertilization and post-fertilization events. These roles are essential in ensuring the organised segregation of nuclei into spores.

1.2.3 Homothallism in *Neurospora* Species

1.2.3.1 The Absence of Strains Containing only Mta

There are homothallic, self-fertile *Neurospora* species where no clear mating type seemed to exist. Homothallic species lack trichogynes and conidia and individuals develop perithecia and go through the sexual cycle without the need for interaction with other individuals. An ascus containing eight self-fertile spores is the end product of a homothallic sexual cycle (Raju, 1977). Following the cloning of mtA and mt**a** of *N. crassa,* Glass *et al* (1990b) were able to probe the genomic DNA from five homothallic species with mtA and mt**a** probes to establish whether these species contained genes with homology to the heterothallic mating type genes. Four out of the five species hybridised to mtA only. The remaining species, *N. terricola,* hybridised with both probes establishing that this species contained both mt**a** and mtA in the same nucleus. Glass *et al* (1990b) isolated seventy new strains of *Gelasinospora* and five new homothallic strains of *Neurospora* from soil samples taken from sites round the world. Thirty five of the *Gelasinospora* strains and all of the *Neurospora* strains were probed with mtA and mt**a**. The *Neurospora* strains hybridised with mtA only and the homothallic *Gelasinospora* strains hybridised with both mtA and mt**a**. No strain was found that hybridised to mt**a** only. Genomic DNA from two homothallic *Sordaria* species (*S. fimicola* and *S. macrospora*), one homothallic *Anixiella* species (*A. sublineata*) and three *Gelasinospora* homothallic species (*G. calospora, G. reticulospora* and *Gelasinospora* S23) were probed with mtA and mt**a** probes and all six species hybridised with both probes.

1.2.3.2 Conservation of A and a Idiomorphs in Homothallic Species

Beatty *et al* (1994) wished to establish to what degree the A and **a** idiomorphs are conserved in homothallic *Neurospora, Anixiella* and *Gelasinospora* species. *N. crassa* probes spanning the entire length of the A and **a** idiomorphs were used to probe homothallic species. A diagram showing the position of the probes in relation to the idiomorphs is shown in figure 1.2.3.2. The mtA probes, A2-A6, hybridised to the *Gelasinospora* species and *A. sublineata* but only A3-A6 hybridised to *N. terricola.* The A2 probe is taken from the 800bp that lie adjacent to the left flank of the mtA idiomorph. A2 includes the mtA-3 ORF, implying that *N.terricola* does not require this gene for functional homothallism. A probe from the right portion of the a idiomorph (**a**4) failed to hybridise with any of the homothallic species. This means the homothallic

species are missing 700bp of the *N. crassa* **a** idiomorph. Probes **a**1-3 and **a**5 hybridised with all the homothallic species tested.

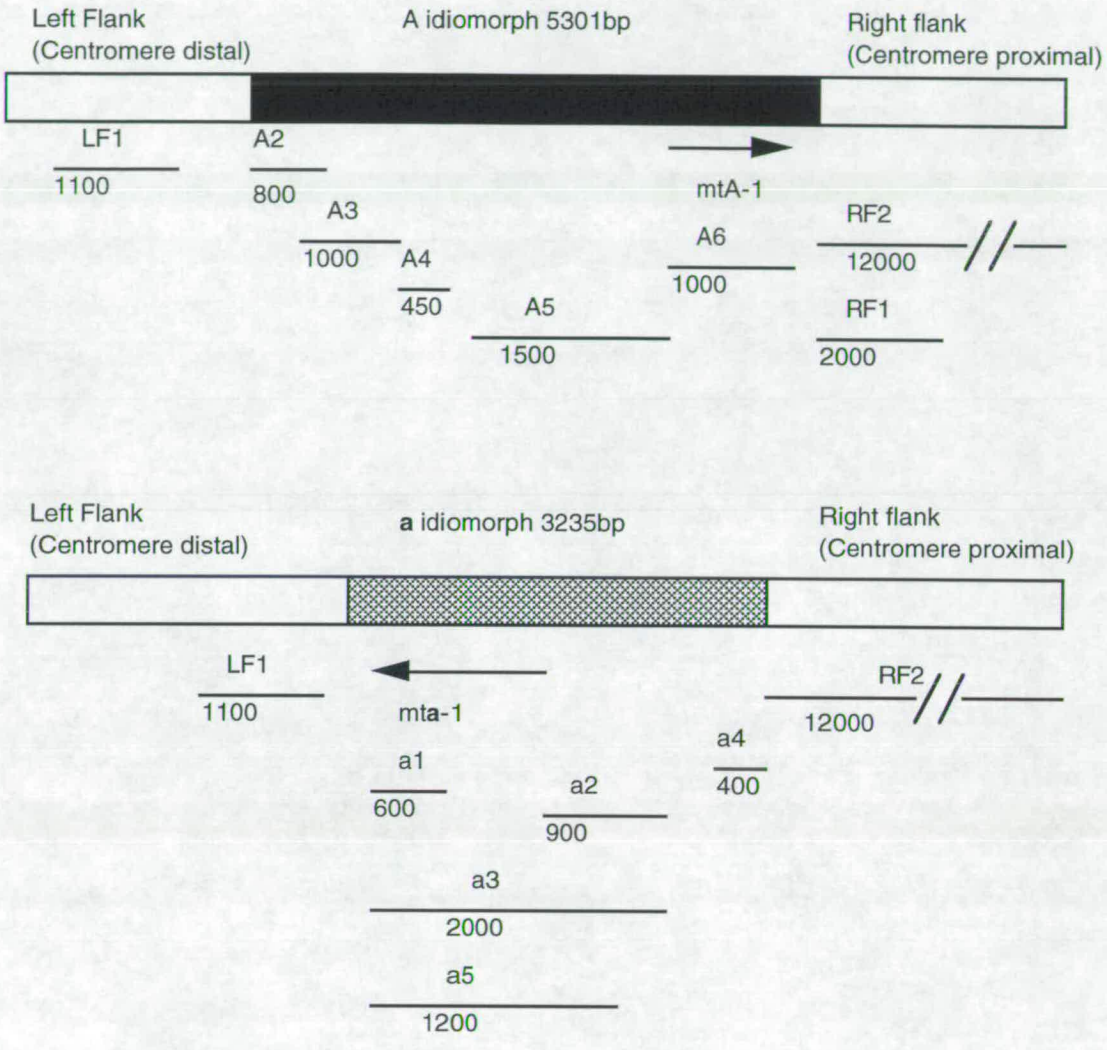


FIGURE 1.2.3.2

The mtA and mt**a** idiomorphs and the clones used as probes with homothallic species. A2-A6 designate probes in the mtA region, **a**1-**a**5 designate probes in the mt**a** region. The mtA-1 and mt**a**-1 open reading frames are shown. RF1, RF2 and LF1 are cloned portions of the centromere proximal and distal regions respectively that flank the mating type locus. The size of each probe in base pairs is shown.

Probes from the right and left flanking DNA of mt**a** and mtA were also used to probe the homothallic species. All the homothallic species hybridised with the probe from the left flank. Hybridisation could not be detected in any of the homothallic species using a 2kb probe from the right flank. The significance of this will be discussed later in section 1.2.4.

The A and **a** probes hybridise to a unique band in OFAGE gels of *N. terricola* chromosomes and this suggests that the A and **a** idiomorphs are present on the same chromosome, leading to the possibility that they are linked. *N. terricola* may then have evolved *via* an unequal crossover event leading to both idiomorphs being adjacent to one another.

It is interesting that mtA-1 and mt**a**-1 remain highly conserved in homothallic species when they are required to bring about fertilisation between individuals of opposite mating type. Homothallic species have no need for this function so perhaps the conservation of mtA-1 and mt**a**-1 reflects their role in post-fertilisation events (Griffiths and Delange, 1978; Griffiths, 1982). One should consider however that fusion of genetically identical nuclei must occur in homothallic species to produce the diploid state. It is possible that mtA-1 and mt**a**-1 remain highly conserved as they are required to bring about this fusion.

1.2.3.3 The Structure and Function of mtA-1 of the Homothallic Species *N. africana*

N. africana is a homothallic species which contains mating type sequences which only hybridise to mtA. Glass and Smith (1994) have cloned, sequenced and analyzed the function of the *N. africana* mtA-1 gene. The probes used by Beatty *et al* (1994), A2-A6, hybridised to *N. africana* genomic DNA establishing that the composition of mtA in *N. africana* is similar to that found in *N. crassa.* The 2kb probe taken from the right flank of *N. crassa* did not hybridise with *N. africana* DNA. The amino acid sequences of the *N. africana* mtA-1 ORF and *N. crassa* mtA-1 showed 88% amino acid identity. The *N. africana* mtA-1 also encoded a region similar to the *S. cerevisiae* MAT α1 proposed DNA binding domain. A single intron 55bp in length is present in *N. africana* mtA-1. The position of the intron is conserved in relation to the position of the intron in *N. crassa.* Glass and Smith (1994) wished to establish if the *N. africana* mtA-1 gene could restore fertility to a sterile *N. crassa* and whether the homothallic mtA-1 would confer homothallism on the recipient mutant *N. crassa.* The mating function of *N.*

africana mtA-1 was assessed by transforming the gene into sterile A^{m64} and $\mathbf{a}^{m1}$ which had lost their vegetative incompatibility function also. The transformants behaved as mtA strains when crossed to mt**a** strains. Successful mating was seen with the formation of perithecia, however ascospore formation was a rare event as this required the transformed mtA-1 gene to directly replace the mutant gene at the mating type locus. As a control the $\mathbf{a}^{m1}$ and A^{m64} mutants were also transformed with vector alone and the transformants used in crosses with A and **a**. No perithecia were observed as a result of these crosses. The mtA-1 transformants were plated individually onto mating medium and observed for signs of self-fertilisation. No perithecia were formed hence the *N. africana* mtA-1 gene did not confer homothallic behaviour on the transformants. The mtA-1 gene did however confer vegetative incompatibility. This is significant bearing in mind that *N. africana* as a species never comes into contact with the mt**a** idiomorph. As *N. africana* evolved, the species retained its mtA mating function and also its vegetative incompatibility function as part of the overall package.

1.2.3.4 Molecular Basis for Homothallism

The observation that *N. africana* mtA-1 does not confer homothallic behaviour on *N. crassa* sterile mutants allows one to conclude that it is not some change to the mtA-1 gene itself which allows *N. africana* to go through the sexual cycle without the need for mt**a**-1. One possibility is that an unidentified second mating type gene exists in *N. africana* which could act in a similar manner to mt**a** but be completely non-homologous to mt**a** (Metzenberg and Glass, 1990). Another possibility could be that the upstream regions of target genes have been altered in *N. africana* allowing these genes to respond to mtA-1 without the need for mt**a**-1. In *N. terricola*, where mtA and mt**a** are present in the same nucleus, perhaps only mtA-1 is expressed in some nuclei and mt**a**-1 in others. This would mean only functionally dissimilar nuclei can fuse making the species functionally heterothallic.

1.2.3.5 Was the Ancestor of *Neurospora* Species Heterothallic or Homothallic?

Metzenberg and Glass (1990) have proposed that heterothallic *Neurospora* species are the most ancient. Pseudohomothallic species exist like *N. tetrasperma* where the asci have four spores instead of eight and each of the four spores contain two nuclei of opposite mating type. Each ascospore gives rise to a self-fertile heterokaryotic culture but

conidia from the self-fertile culture can often produce self-sterile, cross fertile cultures. The first step in the process to homothallism could have resulted in the formation of species like *N. terricola*, where a second mating type idiomorph was acquired, perhaps by aberrant meiosis in a pseudohomothallic strain. The need for the mt**a** idiomorph could have been lost resulting in a species containing mtA only like *N. africana.* Reversing this evolutionary scenario would require the appearance of mt**a** in *N. terricola* and also in a number of related genera, which seems unlikely. One could imagine the situation where enviromental conditions have made it difficult for two individuals of opposite mating type to be brought together. From this homothallism may have been selected for, where species enjoy the benefits of producing resistant spores without the drawbacks of biparental sex. Heterothallic species however have the benefit of genetic variation being introduced as a result of two individuals contributing to the progeny.

1.2.4 Variation in the Flanking DNA of *Neurospora* Species

Randall and Metzenberg (1995) have returned to the issue of the 2kb *N. crassa* probe, taken from the right flank, which does not hybridise with the genomic DNA of homothallic strains originally investigated by Beatty *et al* (1994). Eight probes, the first probe being taken from the centromere-distal or left flank with the rest from the centromere-proximal or right flank, were taken from *N. crassa* flanking DNA. The sequential order of these probes in relation to their position in the flanking DNA is shown in figure 1.2.4a) (taken from Randall and Metzenberg, 1995). Each probe used is defined by its individual fill-pattern. These probes were used to probe the DNA of other heterothallic *Neurospora* species. The species tested are shown in figure 1.2.4a) and where a species lit up with a *N. crassa* probe is seen as a fill-pattern characteristic to the probe. Probe 1 as before was found to hybridise with all the species tested whereas probe 4 showed species-specific and mating type-specific hybridisation. Column I contains "floating boxes" representing *N. crassa* and *N. intermedia* mt**a** probes (shown as probe C). Here no assumption is made about the order of the probes. Some of the *N. crassa* and *N. intermedia* probes did not light up with other species, indicating that a particular species could lack a region of *N. crassa* and *N. intermedia* DNA from its genome entirely. This is shown as the absence of a particular box in column I. Probes from the centromere-proximal flanking DNA of *N. tetrasperma* mtA, *N. intermedia* mtA and *N. discreta* mtA were also used in the hybridisation analysis. Column II shows the hybridisation patterns of probes A and B, taken from *N. discreta* mtA and *N. intermedia* mtA repectively. Column III shows a DNA fragment taken from *N. tetrasperma* mtA which does not hybridise to DNA of any of the other species.

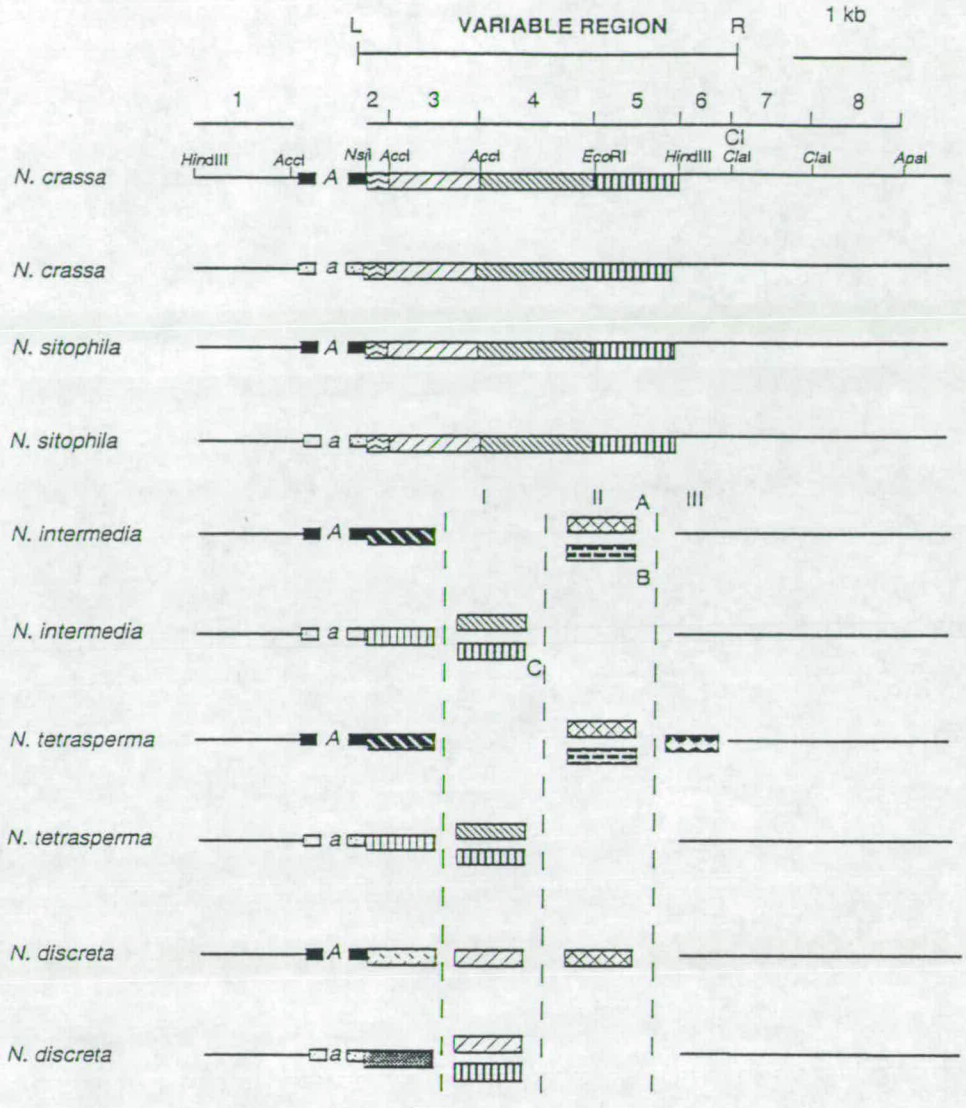


FIGURE 1.2.4a)

Hybridisation differences in the centromere proximal regions of heterothallic *Neurospora* species. The narrow boxed regions, solid black for mtA and dotted for mta represent the idiomorphs. Thin continous lines represent hybridisation with the corresponding *N. crassa* mtA probe. Probes 2-5 are represented by distinct fill-patterns; hybridisation to these *N. crassa* mtA probes is indicated by the presence of the appropriate fill pattern attached to the idiomorph or by a floating box in column I. Floating boxes are represented as such because no presumption is made as to their actual position in the variable region. Column II represents hybridisation with probes from other species (see text for for origin of probes). Column III is a region that is unique to *N. tetrasperma* mtA.

These results show that a 3-5kb region of DNA exists in the centromere-proximal flanks of the *Neurospora* species investigated contains species-specific and/or mating-type specific sequences. These sequences were classed as "variable regions" and were found to be separated from the conserved idiomorph by a "common region", 57-59 bp long which is conserved in all the species examined, both in mtA and mt**a**. The variable region is bordered at its other side by DNA conserved in all the species examined. This is shown as a black line in figure 1.2.4a). A diagram representing the positions of the common and variable regions with respect to the mtA and mt**a** idiomorphs is shown in figure 1.2.4b).

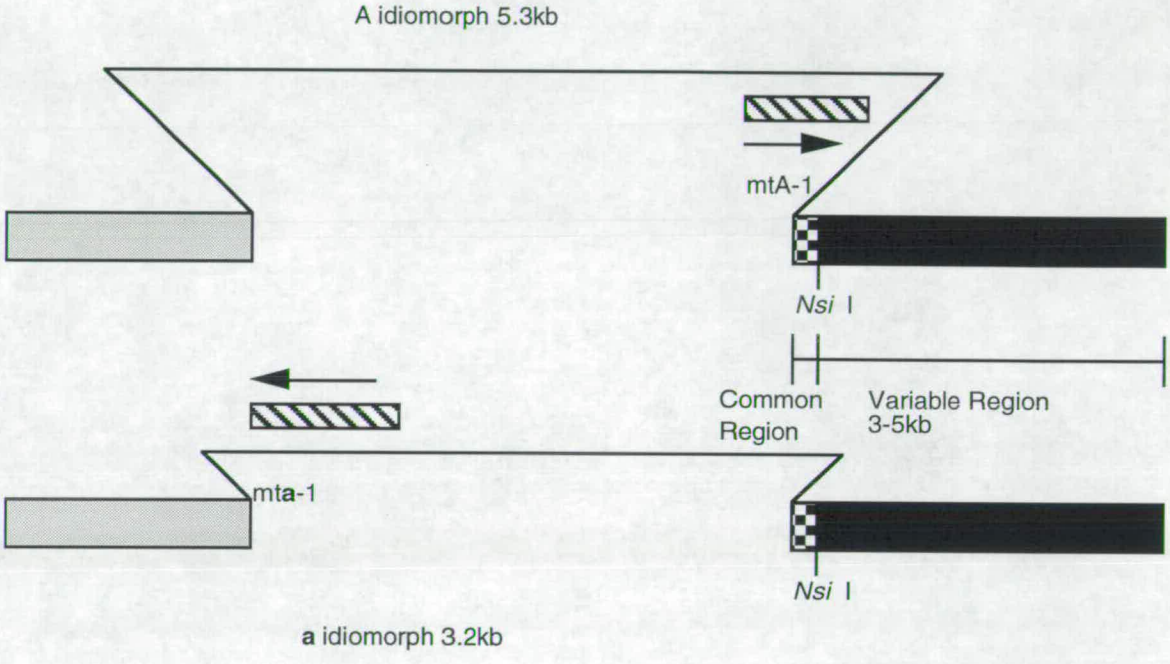


FIGURE 1.2.4b) Schematic diagram of the common and variable regions and their relationship to the mt**a** and mtA idiomorphs. The mtA-1 and mt**a**-1 ORFs are shown as striped boxes. The common regions are shown as checked boxes. The common regions of mtA and mt**a** have a 57-59bp region conserved between both mating types. In mtA the variable region follows directly on from this 57-59bp region but in mt**a** the common region extends for another 80bp making the mt**a** common region approximately 140bp long. The variable regions are shown as black boxes.

The eight probes from *N. crassa* centromere-proximal flank were also used to probe the homothallic species *N. africana, N. lineolata, N. dodgei* and *N. galapagosensis*. These species all contain mtA only. No hybridisation with any of the *N. crassa* probes was found. A 9kb *Eco*RI fragment containing the centromere-proximal flank which did not hybridise to *N. crassa* was cloned from *N. africana* and a portion of it used to probe the other heterothallic and homothallic species. All the homothallic species hybridised with this fragment, none of the heterothallic species showed hybridisation, making it a homothallic-specific variable region.

The first 1kb of the variable region of each of the heterothallic species was sequenced. The mtA variable region and the mt**a** variable region was found to be very similar when compared in *N. crassa* . The same result was found in *N. sitophila.* The other species, *N. discreta, N. intermedia* and *N. tetrasperma* , were very dissimilar in their variable regions when mtA and mt**a** within each species were compared. On comparing the variable region sequences between species, *N. crassa* and *N. sitophila* mtA and mt**a** are very similar to one another. *N. intermedia* and *N. tetrasperma.* mtA variable regions are similar to each other as are their mt**a** variable regions when compared. However" Islands of homology" exist between species over short stretches of their variable regions. For example an island of homology about 300-400bp long exists in the variable regions of *N. africana, N. intermedia* mtA and *N. discreta* mtA. The variable region data suggested that in evolutionary terms *N. crassa* and *N.sitophila* are closely related and form a subgroup. Likewise *N. tetrasperma* and *N. intermedia* appear to be related and form a subgroup. Randall and Metzenberg (1995) discovered that the mtA-1 sequences of these *Neurospora* species gave a similar evolutionary pattern when compared. Figure1.2.4c) shows the tree they obtained using PAUP 3.0, a maximum parsimony method. *N. africana* and *N. discreta* are distant from the more closely related group of *N. crassa, N.sitophila, N. intermedia* and *N. tetrasperma.* The tree does confirm the conclusion from the variable sequence data, that *N. sitophila* and *N. crassa* are more closely related than was previously thought. The same conclusion was also drawn about the relationship between *N. tetrasperma* and *N. intermedia* .

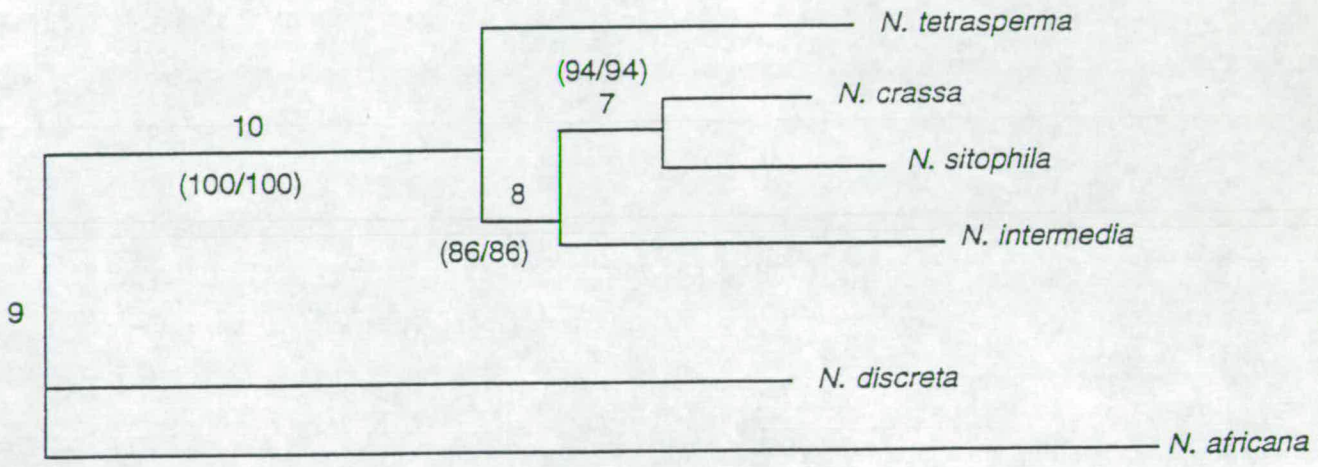


FIGURE 1.2.4c)

Phylogenetic tree compiled using the mtA-1 ORFs of all A mating types. The mtA-1 genes were aligned using PILEUP on the Wisconsin GCG programme. The tree shown was obtained using PAUP version 3 using the ORF and intron for each mtA-1 gene. All characters were unordered and weighted equally. The numbers shown in brackets at the branch points are the confidence levels for bootstrapping of 500 replicates. The branch lengths shown reflect the number of sequence differences.

In the region of the *Acc*I-*Acc*I-*Eco*RI interval shown in figure 1.2.4a) it has been found that A and **a** produce a transcript, differing in size between the two mating types (Randall and Metzenberg, unpublished results). The transcripts increase in concentration under conditions favouring mating. The A variable region transcript was not produced in a strain carrying a frameshift mutation in mtA-1 indicating the transcript is indirectly or directly controlled by mtA-1.

What mechanism accounts for a shift from a highly conserved area (the mtA/mt**a** idiomorph with common region) to a dissimilar variable region? As a Genbank search using the variable sequences did not show the sequences having any homolgy to known transposable or viral elements, then horizontal transfer of transposable or viral elements to account for the variable region seems unlikely. If all the variable regions of different species evolved from a common ancestor then the mutations must have been fixed at a very high rate and also been aimed specifically at the variable regions leaving bordering DNA conserved. So the mechanism by which the variable regions arose must be subject to speculation. A possible function of the variable region could be one involved in speciation, species-specific genes may be found within the variable regions. Randall and Metzenberg (1995) found that there were homothallic specific variable regions which might give an insight into the mechanisms by which homothallic species speciate.

1.3 The Basidiomycetes

Members of the Basidiomycete group produce four basidiospores as products of meiosis, which takes place inside special cells called basidia. Classification of the Basidiomycetes is based mainly on the form of the basidium, one-celled in the Homobasidiomycetes and transversely or longitudinally septate in the Heterobasidiomycetes. The Homobasidiomycetes are divided further into two groups, the Hymenomycetes (which includes *Coprinus cinereus* and *Schizophyllum commune*) and the Gasteromycetes. Similarly the Heterobasidiomycetes are divided into two groups, the Uredinales (rusts) and the Ustilaginales (smuts, which includes *Ustilago maydis*) (Fincham *et al* , 1979).

1.3.1 Life Cycles of the Hymenomycetes and the Ustilaginales

1.3.1.1 *Coprinus cinereus*

Mating in *Coprinus cinereus* is controlled by two unlinked multiallelic loci called A and B (Raper, 1966). For two strains to mate they must differ at both A and B. Figure 1.3.1.1 describes the life-cycle of *C. cinereus.* When a compatible mating takes place the hyphae from each monokaryon fuse, nuclei are exchanged and migrate to the tip cells so apical tip cells contain the resident and migrating nuclei (Buller, 1931). A clamp cell forms on the side of the apical cell and the two nuclei divide. One of the resulting four nuclei becomes trapped in the clamp cell and is released after the clamp cell fuses with the sub-apical cell. Two nuclei from each parent remain in the terminal cell and the nucleus trapped by the clamp cell joins the fourth nucleus in the sub-apical cell. Each cell in the dikaryotic mycelium will contain two nuclei of opposite mating type (Mutasa *et al*,1990). A dikaryon must have different A alleles for formation of the clamp cell (Tymon *et al.*, 1992). Different B alleles allow nuclear migration and clamp cell fusion (Swiezynski and Day, 1960). The two nuclei of the daughter cells fuse and undergo meiosis in a basidium on the undersurface of the mushroom which forms under specific light and temperature conditions (Casselton, 1978).

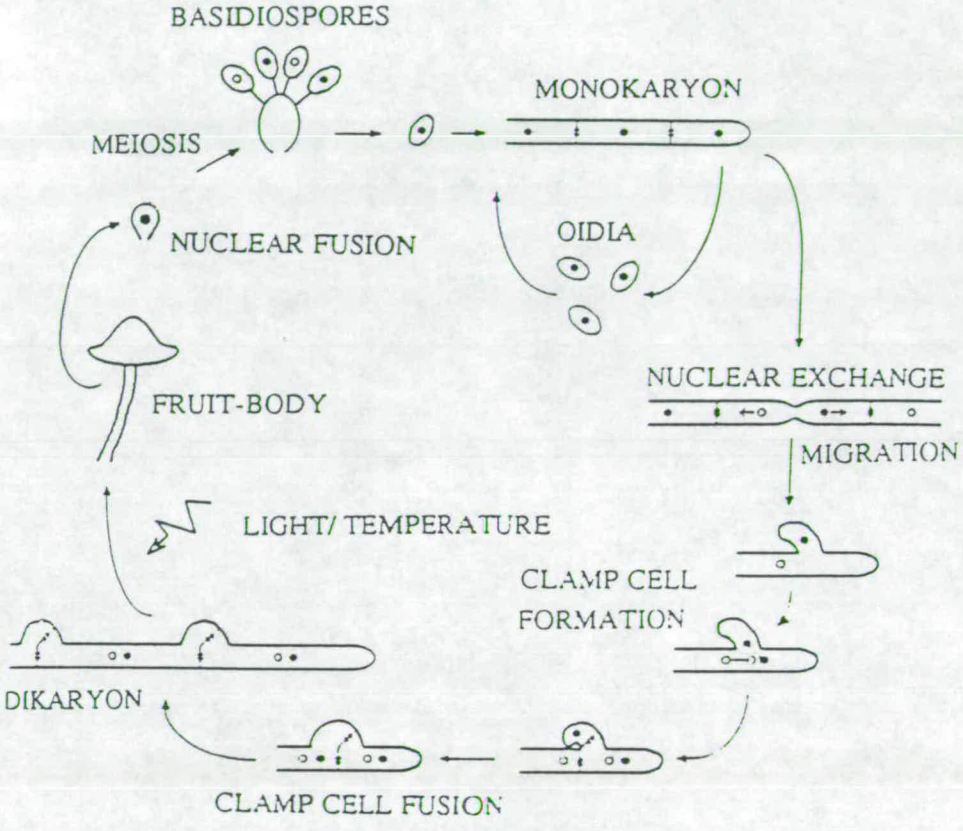


FIGURE 1.3.1.1

Life cycle of *Coprinus cinereus* (Casselton *et al*, 1995)

1.3.1.2 *Ustilago maydis*

Ustilago maydis is a heterothallic corn smut fungus and a parasite in its dikaryotic phase. The life cycle of *U. maydis* is described in figure 1.3.1.2. The life cycle of the organism is under the control of the unlinked **a** and b mating type loci, the two organisms must differ at both **a** and b for a pathogenic filamentous dikaryon to be produced. The **a** mating type locus has only two alleles as opposed to the b locus which has at least 33 different alleles (Holliday, 1961). The **a** locus controls mating through a pheromone recognition system rather like that previously described for *S.cerevisiae* (Bolker *et al.*, 1992). Like mt**a** and mtA of *N. crassa* the **a**1 and **a**2 alleles are completely dissimilar and are flanked by homologous DNA regions but unlike *N. crassa* the **a** locus directly encodes the elements of the pheromone system. **a**1 contains *mfa1*, which encodes the **a**1 mating factor, and *pra1,* which encodes the receptor for the **a**2 mating factor. The **a**2 mating factor is encoded by *mfa2* and *pra2* encodes the receptor for the **a**1 mating factor. For pathogenic development, the dikaryon must carry two different b alleles (Rowell and Devay, 1954, Puhalla, 1970). In the host plant a dikaryotic mycelium develops which produces galls. Karyogamy takes place in the galls to give diploid teliospores (Holliday, 1961). Germination occurs to form a promycelium in which meiosis takes place and haploid basidiospores are budded off.

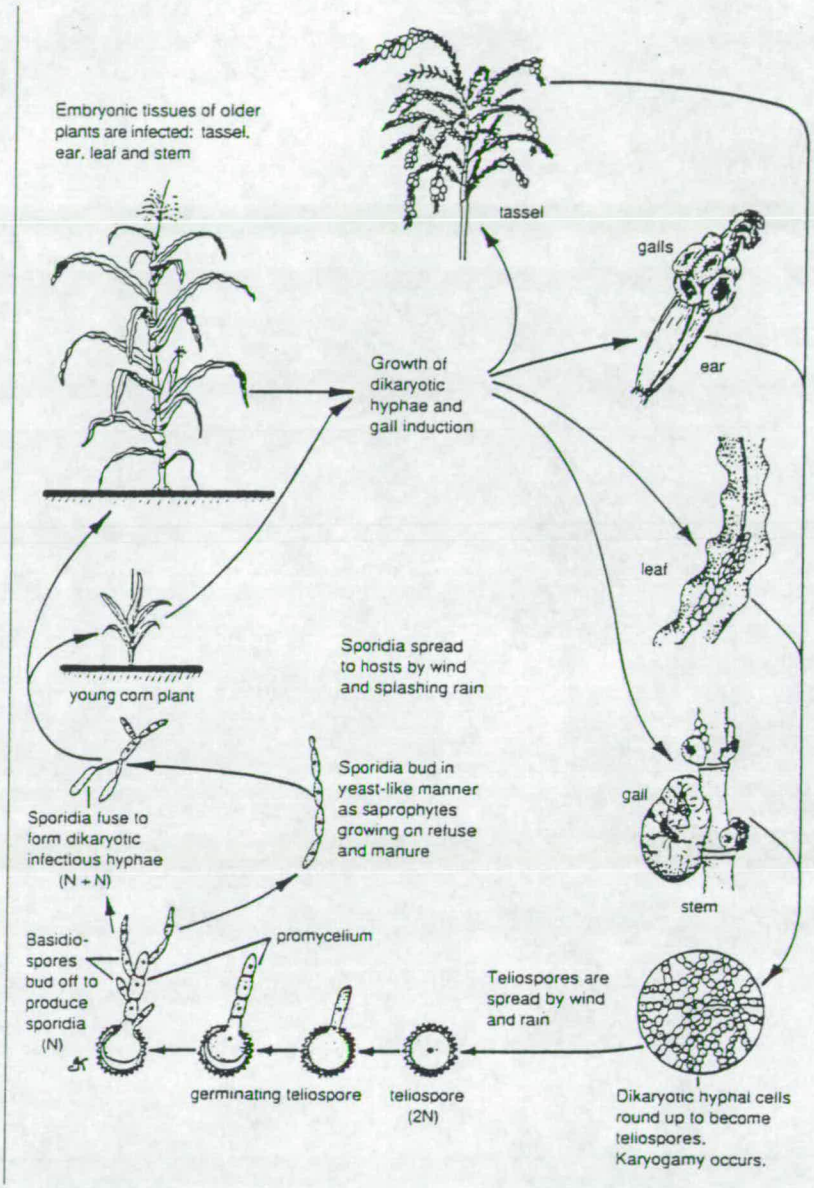


FIGURE 1.3.1.2

The life cycle of *Ustilago maydis* (Banuett, 1992).

1.3.2 Molecular Organisation of the A Mating Type Locus of *Coprinus cinereus* and the b Mating Type Locus of *Ustilago maydis*.

1.3.2.1 *Coprinus cinereus*

Day (1960) demonstrated by recombinational analysis that the *C. cinereus* A locus is composed of linked multiallelic genes called α and β separated by 0.07 map units. About 160 different A-mating types exist (Raper, 1966). For A-regulated development to take place after mating, strains must differ at either α or β at A. The complexity of the A factor has been investigated following the cloning of the A42 mating type factor of *Coprinus cinereus*. The A genes of A42 were cloned by first isolating the *pab-1* gene that is linked to the A mating factor and then initiating a chromosome walk from *pab-1* over the 50kb of DNA between *pab-1* and A42 Strains carrying the A5 (α1β1) or the A6 (α2β2) factors, which contain different A alleles to A42, were transformed with two non-overlapping cosmid clones isolated from the chromosome walk. Both transformations produced unfused clamp cells which indicated that the two non-overlapping clones each contained A genes that were expressed following transformation along with the host A genes to induce A-regulated clamp cell formation. The two clones were proposed to contain the α and β genes (for A42 α3β3). α and β were estimated to be about 1-2kb apart, in contrast to the 7kb estimated by recombinational analysis. Digested genomic DNA from A5 and A6 was probed using A42 probes and no cross hybridisation was found between alleles of these A factors. The A42 factor is embedded in 9kb of unique sequence. This 9kb sequence was found to encode three mRNA's, one of which derived from the region adjacent to the β gene. A possible third gene in the A factor complex was proposed (Mutasa *et al*, 1990).

The A43 mating type factor was cloned from a 35kb cosmid clone by May *et al* (1991). Three non-overlapping fragments were found to produce clamp cell formation following transformation into a recipient strain carrying the A3 mating type factor. It was concluded that the A43 factor was composed of three subunits which were called A(a), A(b) and A(c). All three fragments were used to probe other strains and it was found that some strains failed to show hybridisation to a probe but in contrast strains from different parts of the world would hybridise with the same probe.

A sequence of 40kb from A42, extending into the region flanking the previously discovered two genes, was examined for the presence of new transcripts. Probes covering the 40kb region were used to identify A42 factor-specific mRNAs in Northern

blots of poly (A) RNA isolated from an A42 monokaryon (Kues *et al* , 1992). Seven transcripts were discovered in two clusters of genes, two transcripts in one cluster and the other cluster containing five transcripts. The two gene clusters were separated by 7kb of DNA which produced no transcripts so it was concluded that the two clusters corresponded to the α and β loci identified by classic genetic analysis. The two gene cluster closest to the *pab-1* locus is the α locus and the five-gene cluster the β locus. The α genes were designated α1-1 and α2-1. The five β genes were called β1-1, β2-1, β3-1, β4-1 and β5-1 and it was β1-1, β2-1 and β3-1 that had been identified in the previous study as α and β. None of the genes were found to cross hybridise.

Only four of the A42 genes, α2-1, β1-1, β2-1 and β4-1 were found to induce A-regulated development in transformation experiments using A43, A6, A3 and A5 as hosts. α-2-1 did not produce clamp cell formation in an A6 host and further investigation showed that α2-1 hybridised strongly with A6 DNA indicating that both strains shared the α2-1 allele. Likewise β4-1 did not produce clamp cell formation in the A6 and A43 hosts. Hybridisation analysis showed that A42, A43 and A6 strains share the β4-1 allele. β3-1 appears to have no function in transformation assays. These experiments demonstrate that different strains can have shared α and β alleles but still retain a different overall specificity (Kues *et al* , 1992). The genes of A42 known as α1-1 and β5-1 were found from hybridisation analysis to be shared amongst all A factors tested (A3, A5, A6 and A43). α1-1 was renamed α-fg for α-flanking gene and encodes a metalloendopeptidase, β5-1 was called β-fg for β-flanking gene and was found to encode a protein of unknown function (U. Kues, unpublished data. Quoted in Casselton *et al*, 1995.)

Kues *et al* (1994b) identified three α genes and four β genes in the A43 complex, separated by 7kb of non-coding sequence. Restriction fragments covering the entire A43 factor were used in Northern blots to identify coding regions. The α complex was found to contain α-fg, α1-2 and α2-2. In the β locus β1-2, β2-2, β4-2 and β-fg were identified. The A43 genes were transformed separately into three hosts, A42, A3 and A5, so it could be observed which of these genes elicit clamp cell formation in hosts with different specificities. Only α1-2, β1-2 and β2-2 induced clamp cell formation in all three hosts. Consistent with this result is the lack of hybridisation of these genes to genomic DNAs of the three host strains. The β4-1/4-2 allele is shared between A42 and A43 so failed to induce clamp cell development in an A42 host, but was active in A3 and A5. α2-2 is not shared by the host strains tested yet failed to produce clamp cell

development in all three host strains. The a, b and c fragments identified by May *et al* (1991) contained α1-2, β1-2 and β4-2 respectively.

Kues *et al* (1992) have shown that α2, β1, β2 and β4 encode proteins containing homeodomains, putitive DNA binding domains (Scott *et al*, 1989) . β3-1 encodes a homeodomain which is non-functional in transformation assays. On comparison with *S. cerevisiae* α2 and **a**1 proteins, which also contain homeodomains, β1-1, β3-1 and β4-1 are more similar to α2 and α2-1 and β2-1 are more similar to **a**1 of *S. cerevisiae*. This data suggests that the four genes encode transcription factors. β1-1, β3-1 and β4-1 have been called the HD1 genes and α2-1 and β2-1 the HD2 genes. The HD1 and HD2 genes were found to have transcript sizes of approximately 2.5kb and 2.1kb respectively. In A43 (Kues *et al* , 1994b) α1-2, β1-2 and β4-2 have HD1 homeodomains and α2-2 and β2-2 have HD2 homeodomains, although α2-2 is non-functional in transformation assays. Most homeodomains are made up of three helical structures, the third helical structure being the recognition helix which makes contact with the DNA target site. The other two helices have a stabilizing effect (Kissinger *et al* , 1990). Tymon *et al* (1992) analysed sequence data from β1-1 and showed that unlike most homeodomain motifs, where the most conserved amino acids are WF.N.R found at the recognition helix, in the homeodomain of β1-1 the N is replaced by D which is a conservative change. The N missing from the homeodomain is thought to hydrogen bond to an adenine in every homeodomain-DNA complex (Wolberger *et al* , 1991) so perhaps the homeodomain of β1-1 does not bind DNA strongly. Comparisons to other homeodomain proteins indicates that β1-1 does not seem to have helix II. N-terminal to the β1-1 homeodomain two large helical regions were found which were designated COPsA and COPsB. These helical regions showed 30-35% homology to a POU domain, a DNA binding and dimerization domain of a family of transcription factors (Aurora and Herr, 1992). The POU domain is known to be involved in protein-protein interactions (Ingraham *et al* , 1990) so the COPs domain is a possible dimerization interface. β4-1 also seems to have an unusual homeodomain with a helical region related to the β1-1 COPs domain.

Comparison of the A42 and A43 complexes led Kues and Casselton (1993) to propose the archetypal A gene complex. Two classes of genes, the HD1 and HD2 genes, are found at an A gene complex which can be distinguished by transcript size. In the archetypal gene complex there are four pairs of divergently transcribed genes, each pair containing a HD1 gene and an HD2 gene. These four pairs of genes have been called the a, b, c and d gene pairs (Kues and Casselton, 1992). The A42 α2-1, β1-1, β2-1, β3-1 and β4-1 were renamed a2-1, b1-1, b2-1, c1-1 and d1-1 respectively according to this

nomenclature. Likewise the A43 α1-2, α2-2, β1-2, β2-2 and β4-2 were renamed a1-2, a2-2, b1-2, b2-2 and d1-1. A42 does not contain all the eight genes of an archetypal A complex, the HD2 gene is deleted at the c and d pairs and the HD1 gene at the a pair. A43 does not contain all eight genes either, there is no c gene pair and an HD2 gene has been deleted at the d gene pair. Figure 1.3.2.1 shows a comparison of the A42 and A43 mating-type factors with the archetypal A complex.

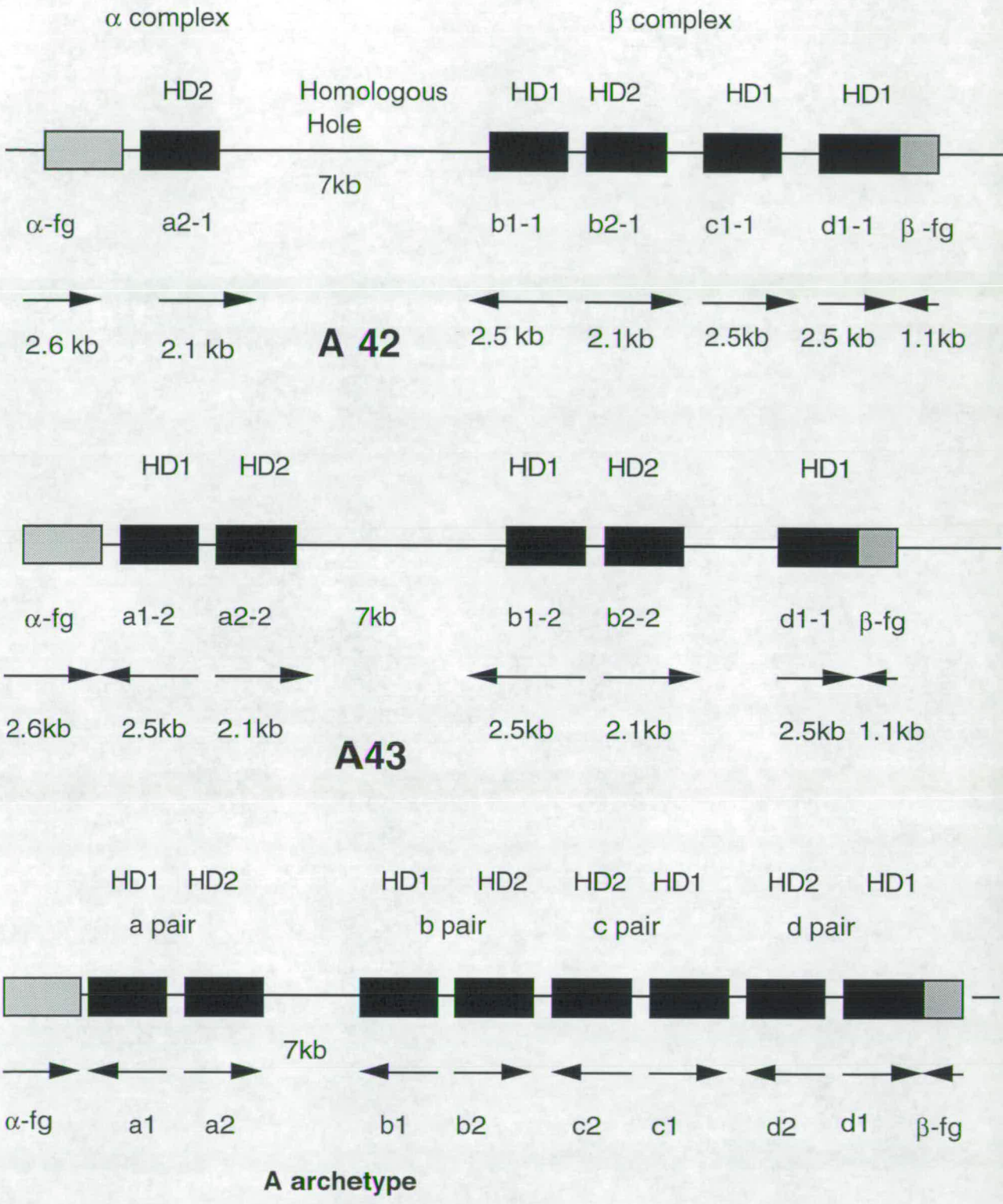


FIGURE 1.3.2.1

Schematic diagram of the A42, A43 and hypothetical A archetypal loci. From the diagram one can see that A42 and A43 do not contain four HD1 and four HD2 genes like the archetypal A locus. Arrows indicate the drection of transcription.(From Kues and Casselton, 1993).

1.3.2.2 *Ustilago maydis*

The b1 and b2 alleles of *U. maydis* were first cloned by Kronstad and Leong (1989). Diploid strains homozygous for b are nonpathogenic and form yeast-like colonies on agar whereas strains heterozygous for b are pathogenic and form mycelial colonies (Day *et al*, 1971). Kronstad and Leong (1989) transformed a cosmid library derived from a b1 strain into a diploid strain homozygous for b2. Transformants carrying the b1 allele should be pathogenic. Transforming activity was found on an 8.5kb *Bam* H1 fragment which was used to probe a b2 cosmid library. Another 8.5kb *Bam* H1 fragment was found to contain the b2 allele. The b3 and b4 alleles were cloned using the b2 fragment to probe genomic libraries (Schulz *et al* , 1990). On sequencing these alleles it was found that all four have a single ORF encoding a putitive polypeptide of 410 amino acids.The first 110 amino acids of the ORF vary between alleles (63% identity) and this was called the variable region. The rest of the polypeptide is highly conserved (90% identity between alleles) and was called the constant region (Shulz *et al* , 1990, Kronstad and Leong, 1990). A homeodomain motif was found in the constant region suggesting that the b proteins have a regulatory role. The homeodomain motif contains the conserved WF.N.R amino acids. The b2 protein sequence also showed that, like β1-1 of *C. cinereus*, b2 has a bipartite DNA binding domain with some homology to POUs at the variable region.

Gillissen *et al* (1992) deleted most of b2 leaving only the N-terminal 10 amino acids and introduced the deletion construct into b1 and b2 strains to determine the functional role of this regulatory protein. Transformants were selected where the resident allele had been replaced by the construct and these transformants crossed with b1 and b2 strains carrying the opposite **a** allele. Both transformants were unexpectedly able to mate as b2 strains. Gillissen *et al* explained this by postulating the presence of a second gene upstream of the b alleles. The known b alleles were renamed bE genes and the new allele called the bW gene. There are as many bW alleles as there are bE alleles. The bW gene has an ORF encoding a protein of 626 amino acids. The bW alleles also have a variable region comprising of the first 130 amino acids (46% identity) with the remaining amino acids making up the contant region (96% identity) encoding a homeodomain. The bE and bW polypeptides are completely dissimilar except for WF.N.R and a few other residues at the homeodomain motif. The organisation of the b locus is shown in figure 1.3.2.2 (Banuett, 1992).

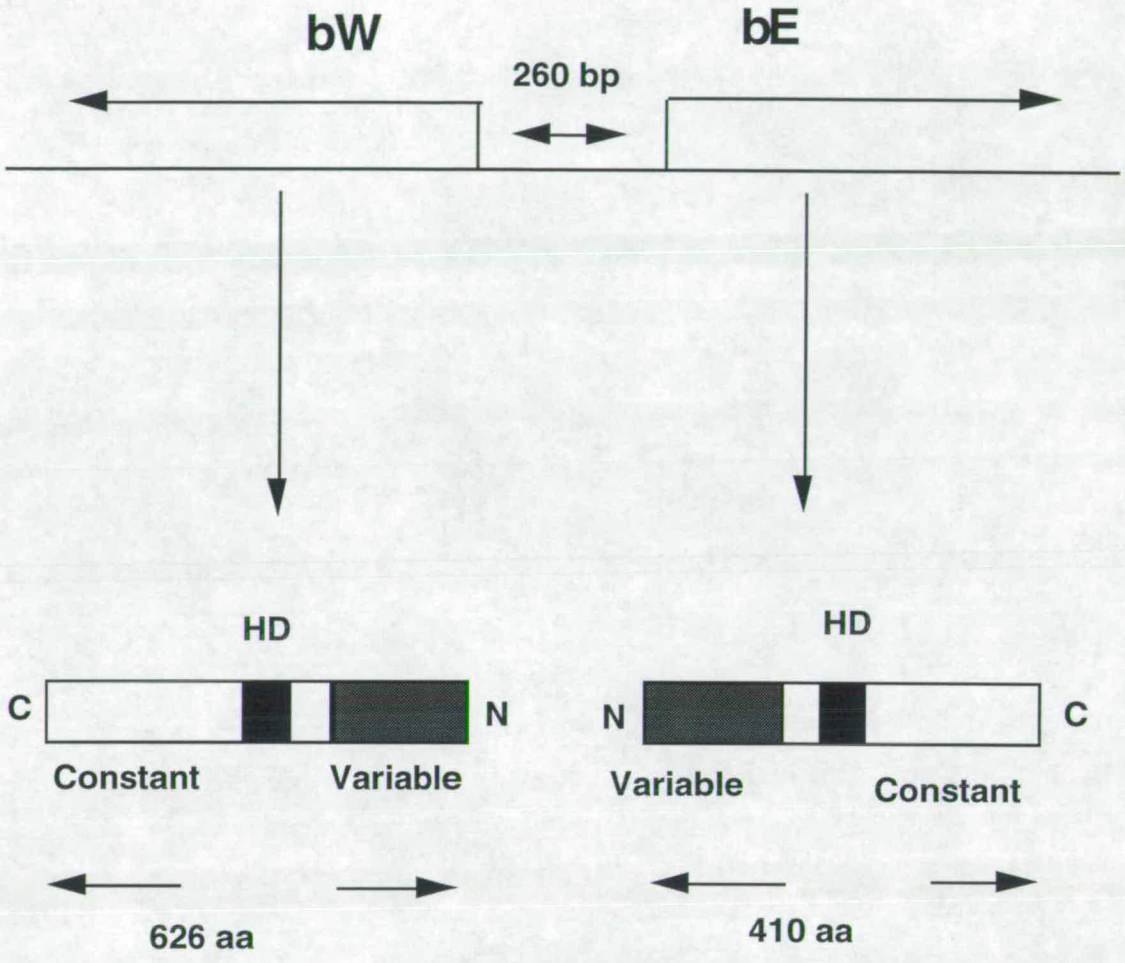
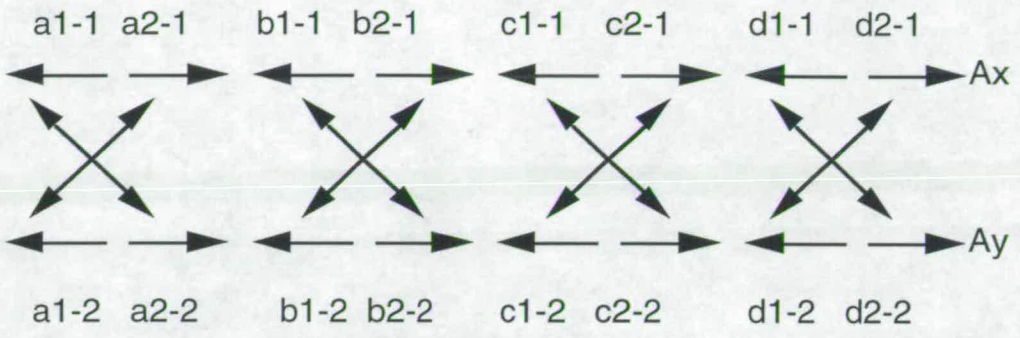


FIGURE 1.3.2.2

Organistaion of the b locus of *U. maydis*. The b locus is comprised of two genes, bW and bE. At the homeodomain in the constant region similarities at the amino acid level are found when the two genes are compared. The rest of the genes are very dissimilar in sequence, but at a structural level bE and bW are similarly organised with constant and variable regions.

1.3.3 Different Alleles of HD1 and HD2 Proteins from the Same Gene Pair at the A locus of *C. cinereus* Form Heterodimers.

The cloning and molecular characterisation of the A locus allows the role of the mating type genes to be interpreted in molecular terms. When the α a1-2 (HD1) gene of A43 is introduced into an A42 host, A-regulated development is induced. This does not happen when the α a2-2 (HD2) gene of A43 is transformed into A42, although this gene is likely to be non-functional (Kues *et al,* 1994b). Classical genetic analysis has shown that the Aα and Aβ genes act independently in inducing A-regulated development (Day, 1960) so it is possible that the single α host gene, the HD2 gene a2-1, is interacting with the A43 α HD1 gene, like the dimerization of **a**1 and α2 of *S. cerevisiae.* Likewise introduction of the A5 α HD1 gene a1-3 into an A42 host generates a compatible HD1-HD2 protein interaction but introducing the A5 α HD2 gene a2-3 into A42 does not promote clamp cell development (unpubl. data quoted in Casselton and Kues, 1994 and Casselton *et al,* 1995). It is probable that the compatible mating here comes from the formation of a heterodimer composed of an α HD1 and an α HD2 protein but for interaction to occur the HD1 and the HD2 proteins have to be from different alleles of the same gene pair. Otherwise mating would not have to take place, the HD1 and the HD2 genes are constitutively expressed (Richardson *et al,* 1993) so heterodimers could form in the monokaryon if self-recognition did not occur. Figure 1.3.3 (Casselton and Kues, 1994) shows the compatible HD1 and HD2 heterodimers that would be formed after a mating between hypothetical Ax and Ay factors. In reality, as no A factor has yet been found to contain all 8 genes, the number of compatible heterodimers will depend on which genes are present and how many have different alleles.



ACTIVE DIMERS

a1-1/a2-2	b1-1/b2-2	c1-1/c2-2	d1-1/d2-2
a1-2/a2-1	b1-2/b2-1	c1-2/c2-1	d1-2/d2-1

FIGURE 1.3.3

Heterodimer formation of mating type proteins after a compatible mating in *C. cinereus*. Only one successful heterodimerization is required to trigger A-regulated development.

Mutations within the A factor have led to to the generation of self compatible alleles, resulting in homokaryons showing A-regulated development. A mutation in A6 (A6*mut*) has given insight into how the need for different alleles of the same gene pair to induce A-regulated development can be overcome. Kues *et al* (1994a) cloned the self-compatible A6*mut* locus by constructing a cosmid library from A6*mut* genomic DNA and probing the library with common flanking sequences from another cloned A locus. Only the self-compatible A6 would be able to promote A-regulated development when transformed into a wild type A6 host. The A6*mut* locus was found to hybridise to both the a2-1 and d1-1 genes in the wild-type A6. These two genes normally are 12kb apart and encode HD2 and HD1 proteins respectively. The A6*mut* locus has a deletion when compared to the wild-type A6 which has brought the a2-1 and d1-1 genes together to encode a fusion protein. The fusion point occurs within a region where a 4bp homology exists between a2-1 and d1-1. The fusion protein only has one homeodomain derived from the HD2 protein. The N-terminal region of the d1-1 protein has been deleted, it is this region that is thought to be the dimerization interface for the HD1 and HD2 proteins (Tymon *et al*, 1992). Yee and Kronstad (1993) have shown that the N-terminal regions of an HD1 protein of *Ustilago maydis* (see section 1.3.5) also determine allele specificity. Tymon *et al* (1992) suggested that even though there is structural homology at the COPs domains when comparing b1-1 and d1-1 for example, the amino acid sequence is variable and therefore might be a way of determining allele specificity. Normally the d1-1 and a2-1 proteins would be identified as self-proteins and dimerization would not occur. The N-terminal region of d1-1 in this case is not required for dimerization or protein-protein recognition. The HD2 protein seems to be sufficient for correct binding activity although based on the *S.cerevisiae* **a**1-α2 model normally the DNA would be bound via both homeodomains. It was interesting to note that a self-incompatible phenotype was restored when the fusion gene was truncated at the 3' end, the C-terminal region of d1-1. This region is a potential activation domain (Tymon *et al* , 1992) and is essential for the fusion protein and hence the heterodimer to function as a transcription factor.

1.3.4 A bE and a bW Protein From Different Alleles can Cause Pathogenic Development in Heterokaryons of *U. maydis*.

It was known that different b alleles were required for pathogenic development. Did this mean that two different bW alleles caused pathogenic development ? Or two different bE alleles ? Or a combination of bE and bW from different alleles ? Gillissen *et al* (1992) took haploid strains with b1 and b2 alleles, bW1 bE1 or bW2 bE2, or with deletion derivatives where only bW1, bW2, bE1 or bE2 were intact and crossed them as shown in Table 1.3.4 Strains were crossed with compatible **a** alleles.

Table 1.3.4 Deletion Analysis of the b locus

Strains	Pathogenic Development
1 bW1 bE1 + bW2 bE2	+
2 bW1 bE1 + bW2 bE2Δ	+
3 bW2 bE2Δ + bW2 bE2Δ	-
4 bW1 bE1Δ + bW2 bE2Δ	-
5 bW1 bE1 + bW2Δ bE2	+
6 bW2Δ bE2 + bW2Δ bE2	-
7 bW1Δ bE1 + bW2Δ bE2	-
8 bW1 bE1Δ + bW2Δ bE2	+

+ indicates filament formation on charcoal medium and tumour induction on corn plants. A Δ sign next to a genotype e.g (bW2- bE2) indicates that particular allele has been deleted (Gillissen *et al*, 1992, Banuett, 1992)

From the table one can see that matings between different bW alleles were unsuccessful in producing pathogenic development, as were matings between different bE alleles. Combinations of bE and bW proteins from different alleles did initiate pathogenic development however. Gillissen *et al* (1992) proposed that, like the HD1 and HD2 proteins in *Coprinus cinereus*, the bE and bW proteins interact to form a heterodimer.

1.3.5 A Region at the N-terminus of bE Genes Determines Allelic Specificity.

Like the HD1 and HD2 proteins from different alleles of the same gene pair in *C. cinereus* , the bE and bW proteins have to be able to distinguish whether they are from the same alleles or different alleles. Yee and Kronstad (1993) identified sequences which determined allelic specificity by constructing chimeric bE alleles. DNA fragments covering a range of deletions in the variable part of the b1E allele (and also carrying the b1W gene inactivated by the selectable marker hygromycin B) were transformed into a b2E strain. Targeted gene replacement occurs at a high frequency in *U. maydis* so chimeras between b1E and b2E were created. Recombination takes place within bW but the recombinant gene is inactive due to the hygromycin B marker. Three classes of transformants carrying a chimeric gene were determined from whether the transformant produced a white, ariel mycelium (compatible reaction) or a flat nonmycelial colony (incompatible reaction) when mixed with a1b1 or a1b2 strains. Class I transformants had retained the ability to mate as a b2E strain as indicated by the incompatibility reaction with a1b1; that is the intergration into the b2E gene had not interrupted the region determining specificity. Class II transformants were able to mate with both b1 and b2 strains. Class III transformants behaved as a b1E strain, mating with the a1b2 strain. PCR products from the transformants were sequenced to determine the positions of the recombination events that accounted for the phenotype of each transformant. Class I transformants had recombination points in the region between codons 28 and 39. Recombination at codons 48, 49, 51, 60, 70 and 79 gave alleles that differed in specificity to b1E and b2E so that Class II transformants could mate with b1 and b2 strains. Class III transformants had recombination points in the region between codons 87 and 156. The results indicated that a region between codons 39 and 87 determines allelic specificity. One could predict that recombination during meiosis in this region could result in alleles with altered specificity when compared to the parental specificities. So a domain that determines allele specificity is found in the N-terminal portion of the bE gene of *U. maydis*.

1.3.6 The 5' Ends of A genes Determine Gene Specificity

Kues *et al* (1994c) confirmed that the variable 5' ends of *C. cinereus* A genes determine gene specificity.The 5' ends of of A42 genes were exchanged so that two chimeric genes were generated, an HD2 gene with the 5' end of b2-1 and the 3' end of a2-1, and an HD1 gene with the 5' end of b1-1 and the 3' end of d1-1. The chimeric genes were introduced into A42, A6 and A5 hosts. A42 and A6 share a2-1 and d1-1 but have different alleles of b1 and b2, hence if the 5' ends of the genes determine specificity then the chimeric genes will cause clamp cell development when transformed into an A6 host. A5 and A42 do not share any specificity genes so all four A42 genes should promote A-regulated development in an A5 host. The chimeric genes failed to promote clamp cell development in an A42 host showing no new specificity had been created. The chimeric genes were functional as both initiated A-regulated development in the A5 host. Both chimeric genes behaved as b gene alleles in the A6 host indicating that the 5' ends of the genes determined gene specificity. The chimeric genes were not identified as new a2-1 or d1-1 alleles. These experiments did not show that the N-terminal regions determined allele specificity.

1.3.7 Protein Structure of Four A42 Gene Products

In section 1.3.2.1 the protein structure of b1-1 (β1-1) was mentioned, its unusual homeodomain with D replacing the N in the recognition helix and the presence of helical COP1 and COP2 regions N-terminal to the homeodomain, potential dimerization interfaces (Tymon *et al* , 1992). Kues *et al* (1994 c)) examined the complete sequences of the four A42 specificity genes and predicted from these which regions are of functional significance when HD1 and HD2 interact. b1-1 and d1-1 are HD1 genes, a2-1 and b2-1 HD2 genes as mentioned before. d1-1 was found to have a similar predicted protein structure when compared to b1-1, two α-helical regions were proposed to be N-terminal to the homeodomain and another helical region was found C-terminal to the homeodomain. This helical region is similar to the short charged C-terminal tail of α2 of *S. cerevisiae* which is proposed to play a part in dimerization with a1 (Mak and Johnson, 1993). There is another predicted helical region at the C-terminus of d1-1, but this can be deleted and have no affect on promoting A-regulated development *in vivo* (Tymon *et al* , 1992). d1-1 was found to have the atypical WF.D.R in the homeodomain region.

The HD2 genes, a2-1 and b2-1, were found to have homeodomains conserved for WF.N.R. The predicted helical structures on either side of the homeodomains were present too in the HD2 proteins. The C-terminus in HD2 proteins was found to have an overall positive charge due to a high lysine and arginine content. This is unlike the C-terminal ends of HD1 proteins which are rich in serine and threonine and partially proline rich and negatively charged near the C-terminus. The C-terminal regions of HD1 proteins have been proposed to play a role in activation (Tymon *et al* , 1992, Kues *et al* , 1994a). Kues *et al* (1994c) found that much of the 3' end of HD2 genes can be deleted without loss of function. Regions close to the 3' ends of HD1 genes are not dispensable, the minimal functional HD1 protein contained 594 amino acids out of a possible 632 in b1-1.

Figure 1.3.7 shows the predicted organisation of HD1 and HD2 proteins (Kues *et al*, 1994c)

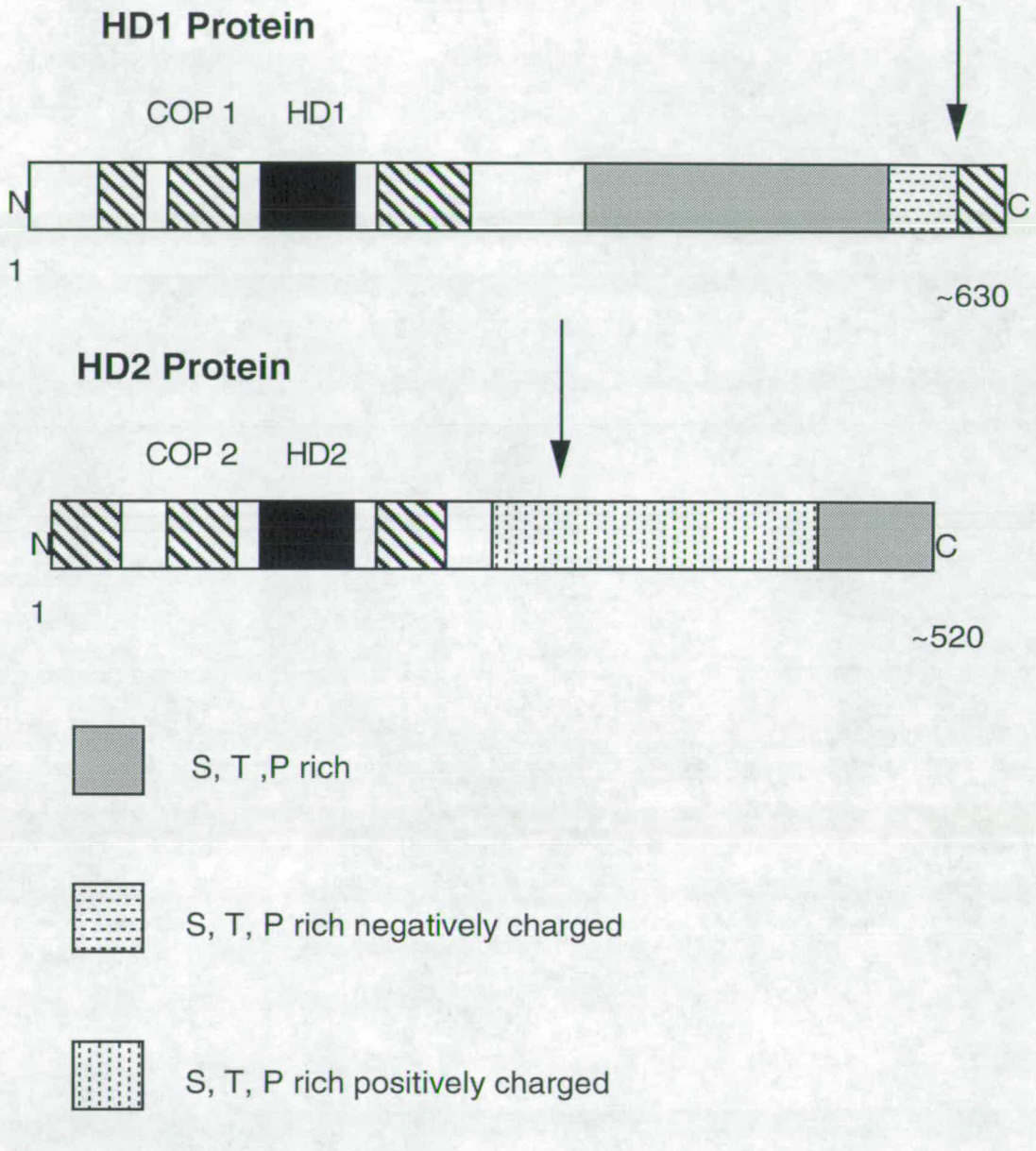


FIGURE 1.3.7

Representation of the predicted structure of the HD1 and HD2 A proteins of *C. cinereus*. The homeodomains are shown as black boxes and the helical regions shown as striped boxes. Other patterns in the figure refer to the characteristics of the c-terminal regions. These regions were truncated without loss of function as far as the arrows indicate. Numbers on the diagram refer to the amino acids.

1.3.8 HD1 and HD2 Proteins of *C. cinereus* Dimerize Via an N-Terminal Domain.

Banham *et al* (1995) determined that the N-terminal domain of an HD2 protein was not required for regulating gene transcription in the A6 mutant fused heterodimer. As mentioned the fusion protein lacks the N-terminal region of the HD1 protein, including the homeodomain, hence this region is not essential for regulating transcription of genes targeted by the mutant fusion protein. 5' deletions were made to the HD2 part of the fusion protein, the homeodomain was left intact. Two fusion proteins with HD2 5' deletions, amino acids 2-61 and amino acids 2-143, were transformed into an A6 host and both transformations resulted in the formation of clamp cells. It was concluded that the N-terminal domain of the HD2 portion of the fused heterodimer was not required for constitutive activity of the fused heterodimer.

The most likely function of the N-terminal domains of HD2 and HD1 proteins other than to determine allele specificity is to form the dimerization domains. The COP domains are found at the N-terminal regions of HD1 and HD2 proteins, these are the putitive dimerization domains. Banham *et al* (1995) used the *in vitro* glutathione S-transferase (GST) association assay to show dimerization between A proteins. Six different proteins from A6 and A42 were expressed in *Escherichia coli* as GST fusion proteins. The fusion proteins were bound to glutathione S-Sepharose beads and full length HD1 proteins , ^{35}S labeled by translation *in vitro*, adsorbed to the bound fusion proteins. Bound HD2 fusion proteins contained the entire N-terminal region and the homeodomain, with varying amounts of C-terminal sequences. The A6 HD2 b2-3 fusion protein retained its compatible HD1 partner from A42, b1-1. None of the HD2 fusion proteins from A42 retained b1-1. Likewise the A42 HD2 b2-1 fusion protein retained its compatible HD1 partner from A6, b1-3. The 5' ends of the b1-1 and b1-3 HD1 genes were exchanged and the *in vitro* glutathione S-transferase assay repeated as before. The b1-1:b1-3 protein was retained by b2-3 and the b1-3:b1-1 protein retained by b2-1, showing that the area that determines whether a compatible dimerization takes place lies within the region where the the 5' ends were exchanged, that is within the first 158 amino acids of HD1 proteins. This experiment also demonstrated that the 5' ends of A genes determine allelic specificity. The first 163 amino acids of b1-1 and the first 63 amino acids of b1-3 were sufficient for dimerization *in vitro* with their compatible HD2 partners. The truncated b1-3 protein did not contain the homeodomain.

Incompatible proteins are distinguished from compatible proteins in that they are unable to heterodimerize. COP coiled-coil motifs are found at the N-termini of A proteins and it is likely that variable amino acids at these motifs determine whether two proteins will dimerize.

1.3.9 bE and bW Dimerize if They are From Different Alleles

Using the two-hybrid system, Kamper *et al* (1995) have shown that the bE and bW proteins from different alleles dimerize. The coding regions of bW1, bW2, bE1 and bE2 were fused to either the C terminus of the GAL4 DNA binding domain (GB) or the N terminus of the GAL4 DNA activation domain (GA). Different combinations of the GB and GA fusion proteins were transformed into *S. cerevisiae* strain Y153 with the expectation that if the fusion plasmids were compatible, dimerization would take place. This would be seen as β-Galactosidase activity. bE and bW proteins from different alleles were found to interact, for example pGB-E1 combined with pW2-GA could activate transcription of the *lacZ* reporter gene. bE and bW proteins from the same alleles did not interact in the two-hybrid system. Recognition of self/non-self functions at the level of protein-protein interaction. C-terminal deletion derivatives of the bE and bW proteins were fused to either GA or GB and tested in the two-hybrid system. The N-terminal fragments of the variable domains showed differing levels of interaction, shown as β-Galactosidase activity, or lost the ability to interact altogether depending on the the length of the fragment and allelic combination. Contacts between the variable domains of bE and bW are necessary for heterodimer formation. The homeodomains can be deleted without interfering with heterodimer formation.

The variable region of bE2 was mutated using misincorporating polymerase chain reaction. Several different point mutations of bE2 were recovered which were able to form heterodimers with bW2 in the two-hybrid system. Plasmids carrying the mutant bE2 alleles were transformed into a b2 strain and tumours were induced on corn plants after 5-8 days indicating that the mutant alleles were functioning in combination with bW2.

Five single point mutations in bE2, allowing bE2 to function in combination with bW2, resulted in an increase of hydrophobicity. Kamper *et al* (1995) proposed a model for bE and bW interaction that takes this and the data from Yee and Kronstad (1993) into account. Interactions between bE and bW proteins may occur through hydrophobic as

well as polar interactions. N-terminal fragments of the variable region show different levels of interaction depending on their length indicating that there is more than one contact point. An N-terminal fragment of bE1 and bE2 (consisting of the first 77 amino acids) can be of the same size yet in the two hybrid system only the truncated bE1/bW2 combination was active. This indicates that contact sites are located in different positions in different allelic combinations. Failure to form heterodimers in allelic interactions results from having interfering amino acids at corresponding contact sites in bE and bW. In non-allelic combinations an interfering amino acid in one protein would not be opposed to an interfering amino acid in the other protein so there is no repulsion and heterodimerization can occur. For dimerization not to occur, several amino acids at corresponding contact sites in both proteins have to interfere. A single amino acid change in bE2 created a hydrophobic interaction that allowed heterodimerization with bW2. The behaviour of the chimeric bE1/bE2 polypeptides (Yee and Kronstad, 1993) can be explained by the model too. The chimeric bE1/bE2 polypeptide that behaves like bE2 does so because all its critical contacts are provided by the bE2 portion of the protein so the chimera will only interact with bW1. Interaction with corresponding contact sites in the bW2 protein results in repusion between the chimera and bW2. The bE1/bE2 chimera with new specificity has a fusion point between the critical cohesive residues in bE1 and bE2 so the chimeric polypeptide can interact with bW1 and bW2.

1.4 Scope of Thesis

λ clones containing putative *Sordaria* mating type genes were available for use in this project. As these clones had been isolated with *N. crassa* probes it was of interest to see how closely related *Sordaria* and *Neurospora* species are to each other. The aims of the project were as follows:

1 To subclone and sequence as many *Sordaria* mating type genes as time permitted.

2 To express a homothallic *S. equina* mtA-1 gene in a sterile heterothallic *Neurospora* mutant. Could the *Sordaria* mtA-1 gene restore mating type and vegetative incompatibility function on the mutant? Could the *S. equina* mtA-1 gene confer homothallism on the recipient mutant?

3 To construct a phylogeny to establish the relationships between *Sordaria* and *Neurospora* species.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Solutions for DNA work.

TBE 10X stock solution

108g Tris Base
55g Boric acid
9.5g EDTA
distilled water to 1 litre

TE

10mM Tris-HCL pH 8.0
1mM EDTA pH 8.0

6X Loading buffer

1.75g Ficoll
1ml 1M Tris pH8
2ml 0.5M EDTA
0.1g Bromophenol blue
distilled water to 10ml.

50X Denhart's reagent

5g Ficoll
5g Polyvinylpyrrolidone
5g Bovine serum albumin
distilled water to 500ml filtered, diluted to 5X for use.

Hybridisation solution

1mM EDTA
0.5M NaP,
(1M Na_2HP0_4, 72ml; 1M NaH_2P0_4, 28ml)
7% SDS

Wash Solution

40mM NaP (as for hybridisation solution)
1% SDS
1mM EDTA

Spermidine-SDS buffer	4mM spermidine 10mM EDTA 0.1M NaCl 0.5% SDS 10mM ß-mercaptoethanol 40mM Tris HCl pH 8.0
Extraction buffer for use in fungal DNA extractions	1% CTAB 0.7M NaCl 50mM Tris pH 8 10mM EDTA 1% β–mercaptoethanol
STET	8% Sucrose 0.5% TritonX 100 50mM EDTA pH 8.0 50mM tris HCl pH 8.0
GTE buffer (used for alkaline lysis method of preparing plasmid DNA)	50mM glucose 25mM Tris pH8 10mM EDTA pH8
Denaturing solution	2M NaCl 0.5M NaOH
Neutralising solution	1M Tris HCl pH 7.5 3M NaCl
SSC 20X Stock solution.	175g NaCl 88g Sodium Citrate adjusted to pH 7.0 distilled water to 1litre

2.1.2 Media for growth and manipulation of bacteria and phage.

Phage Buffer	0.3% KH_2HPO_4 0.7% Na_2HPO_4 0.5% NaCl 1mM $MgSO_4$ 0.1mM $CaCl_2$ 0.11% gelatin solution
Luria-Bertoni medium (LB)	1% Bacto-tryptone 0.5% Bacto-yeast extract 1% NaCl adjusted to pH 7.2 with NaOH
LB agar	LB medium supplemented with 1.5 % agar
BBl Top agar	1% Baltimore Biological Laboratories trypticase 0.5% NaCl 0.65% agar
LB-amp	LB supplemented with ampicillin (50µg/ml)

2.1.3 Media for growing, crossing and transforming fungi.

Vogels growth medium	1.5% Difco Bacto Agar 2% glucose 1X Vogels salt solution
Growth medium for germinating conidia	1X Vogels N salt solution 1.5% sucrose 0.1mg/ml adenine

Underlay agar	0.05% glucose 0.05% fructose 2% sorbose 2% agar (Difco Bacto) 1X Vogels salt solution (Nitogen free where selecting for Ignite resistance) 0.1mg/ml adenine (0.5% proline as a nitrogen source where selecting for Ignite resistance) (Ignite at a concentration of 200µg/µl for selection plates)
Overlay agar for checking for spheroplasts	As for non-selective underlay except bacteriological agar was used instead of Difco Bacto agar.
Overlay agar for viability plates	As for non-selection underlay except 1M Sorbitol was added to provide osmotic protection and bacteriological agar was used instead of Difco Bacto agar.
Growth medium for regeneration of spheroplast cell walls	0.05% glucose 0.05% fructose 1 X Vogels salt solution (N-free) 1M Sorbitol 0.5% proline 0.1mg/ml adenine
Medium for selective slants	2% sucrose 2% Difco bacto agar 0.5% proline 1X Vogels salt solution (N-free) 0.1mg/ml adenine Ignite at 200µg/µl

Crossing medium	2% sucrose 2% Difco bacto agar Westergaard and Mitchell salt solution 0.1/mg/ml adenine

2.1.4 Bacterial Strains, Bacteriophage and Plasmid Vectors

E-coli strains used in the course of this project.

E-coli strain	Genotype	Reference
Q358	*supE hsdR* ϕ 80^{r}	Maniatis *et al* (1989)
XL-1 Blue	*recA* 1, *endA* 1, *gyrA* 46,*thi* , *hsdR* 17, *supE* 44, *relA* 1, *lac* , [F'*proAB lacI* ꟾZΔ*M15* Tn*10* (Tet^{r})]	Bullock *et al* (1987)
JM109	*recA* 1, *endA* 1, *gyrA* 96, *hsdR* 17, (r_{k-},m_{k}+), *telA* 1, *supE* 44,λ^{-}, *thi* , Δ(*lac-proAB*), [F, *traD* 36, *proAB,lacI* ꟾ ZΔ*M15*]	Yanisch-Perron*et al* (1985)

Bacteriophage	Reference
EMBL3	Frischauf *et al.* (1983)
EMBL4	Frischauf *et al.* (1983)

Plasmids used in the course of this project

Plasmid	Reference
pTZ18R	Pharmacia P-L Biochemicals Molecular and Cell Biology Catalogue
p Bluescript SK+/KS+	Stratagene Cloning Systems Catalogue
pBARGRG1	Pall and Brunelli (1994)
pGEM®-T vector	Promega Technical Bulletin No 150.

2.2 Methods

2.2.1 Manipulations of Bacteria and Phage

2.2.1.1 Plating Cells

A single colony from a Q358 streaked plate was picked and grown up overnight, shaking at 37°C, in 10ml of L-Broth. The overnight culture was then diluted 1:10 in L-Broth + 0.2% maltose and grown for 1 hour. The culture was spun down at 3000 rpm for 10 minutes and the supernatant poured off. 5ml of sterile 10mM $MgCl_2$ were added and the cells resuspended ready for infection with lambda.

Maltose improves the efficiency of bacteriphage adsorption as it induces the lambda receptor (lamB protein) and Mg^{2+} ions play an important part in the maintenance and integrity of lambda phage particles and aid phage adsorption.

2.2.1.2 Competent Cells

A single colony from a plate streaked with XL-1 Blue was picked and grown up overnight, shaking at 37°C, in 10mls of L-Broth. The overnight culture was diluted 1:20 with L-Broth and grown to a OD_{550}=0.5. The cells were spun down at 3000 rpm for 10 minutes and resuspended in 10mls of sterile 50mM $CaCl_2$. The cells were left on ice for 20 minutes and spun down again. This time the cells were resuspended in 2mls of $CaCl_2$. The cells were then ready to be used in transformations.

2.2.1.3 Transformations

100µl of XL-1 Blue competent cells with 2µl of ligated DNA (concentration 0.1-10ng) were placed on ice for 30 minutes. The cells were heat shocked for 2 minutes at 42°C and placed back on ice for 90 seconds. 50µl of L-Broth were added to the cells and the mixture incubated at 37°C for 30-60 minutes. The cells were then plated out onto L-Amp plates and incubated at 37°C overnight.

2.2.1.4 *Sordaria* mating type gene clones

Genomic libraries were available for several *Sordaria* species in either cloning vector lambda EMBL3 or EMBL4 (Frischauf *et al* , 1983). Fragments between 8 and 21kb in

length can be cloned in these vectors. The desired recombinant phage were identified using a radiolabelled probe (Benton and Davis, 1977), the probes used here being subclones of the *Neurospora crassa* mating type genes mtA and mt**a**. All *Sordaria* lambda clones were produced as part of the Genetics Honours cloning practical.

2.2.1.5 Plating out lambda phage

100µl of Q358 plating cells were transferred to a test tube. Depending on the titre of the phage lysate, 10µl of a particular dilution (the phage diluted in phage suspension buffer) was added to the plating cells and the tube was incubated at 37°C for 20 minutes to allow for attachment of phage particles. While waiting a 3ml aliquot of BBL top molten agar was transferred to a test tube in a 55°C waterbath. After 20 minutes, the top agar tube was removed from the waterbath, flamed, and the agar poured into the tube containing the infected plating cells. The contents of this tube were then poured over a dried L-Broth plate and the top agar allowed to set. The plate was incubated at 37°C overnight.

2.2.1.6 Lambda lysates

To obtain a 500ml lysate from which lambda DNA can be extracted, the Q358 cells must be infected with phage at a titre of 10^{10} pfu. Plate lysates were made to obtain this high titre. 5ml of phage suspension buffer were poured over an L-plate where confluent lysis of the plate had occurred after plating out 10^5 pfu of phage. After 4 hours of gentle shaking the 5ml of phage suspension buffer were poured into a Falcon tube and the tube was spun to pellet any debris. The supernatant was removed and 100µl of chloroform added to free any phage from intact cells. After another spin the supernatant was recovered and a drop of chloroform added. The lysate was stored at 4°C. The lysate was approximately 10^{10} pfu/ml. 1ml of overnight Q358 cells were then added to 500ml of pre-warmed L-Broth + 0.2% maltose. The cells were grown to an OD_{600}=0.5. The flasks were then innoculated with 1ml of plate lysate and 10mM $MgCl_2$ also added. The culture was shaken at 37°C till lysis occurred which is apparent by the presence of cell debris. Chloroform was added to the lysate to a final concentration of 1% and the lysate spun down at 4°C for 10 minutes at 10000 rpm. The supernatant was then removed ready for DNA extraction.

2.2.2 DNA Manipulation, Preparation and Detection

2.2.2.1 Phenol/Chloroform Extractions and Ethanol Precipitations

DNA was usually purified with phenol, followed by phenol/chloroform and finally chloroform on its own. To the DNA in TE an equal volume of phenol, equilibrated against TE, was added and the solution mixed and spun down in a microcentrifuge briefly. The upper band containing the purified DNA was removed to a fresh Eppendorf. To the DNA an equal volume of 1:1 phenol/chloroform solution was added and the procedure repeated as before. Finally the DNA was extracted using an equal volume of chloroform. This final stage allows the removal of any phenol residues. The DNA was precipitated by adding 0.1 volume of 3M Na Acetate and 2 volumes of absolute ethanol. The Eppendorf was incubated at -70°C for 10 minutes followed by a 10 minute spin in the microcentifuge. The supernatant was removed and the DNA pellet rinsed in 70% ethanol, dried under vacuum and resuspended in TE.

2.2.2.2 Agarose Gel Electrophoresis

Agarose (0.7-1%) gels were used to analyse restriction digests of DNA and the integrity of the DNA. Gels were prepared and run in 1X TBE buffer with ethidium bromide included in the gel at a concentration of 0.5µg/ml. The sample of DNA was loaded in 1X stop buffer into the wells of the submerged gel. Gels were generally run at 90V and the DNA visualised using short wave-length UV light.

2.2.2.3 DNA Restriction.

Restriction enzymes were supplied by Boehringer Mannheim. DNA was digested at 37°C using the buffer appropriate to the enzyme. Genomic DNA was left to digest overnight, lambda DNA for 3-4 hours and plasmid miniprep DNA for 1-2 hours.

2.2.2.4 Extraction of DNA from agarose

The desired DNA fragment was cut out of a low melting point agarose gel using long-wave UV light and an equal volume of 1X TBE/0.2M NaCl was added to the gel slice. After melting the agarose for 10 minutes at 65°C a 3/4 volume of phenol buffered in 1X

TBE/0.1M NaCl was used to extract the DNA from the agarose. Two phenol extractions were required in total and if the DNA remained in a volume greater than 400µl, butanol was used to reduce the volume. The DNA was ethanol precipitated and resuspended in 10µl of TE.

2.2.2.5 Ligations

DNA fragments were ligated into the vector of choice using T4 DNA ligase supplied by Boehringer Mannheim. The DNA fragment and digested vector have complementary overhanging ends. 40ng of digested vector were added to the DNA fragment, which should be in excess concentration, along with 10X ligase buffer (1µl in a 10µl total) and the T4 DNA ligase (1µl, 1U/µl). The ligation mix was left at 16°C overnight.

2.2.2.6 DNA preparations

2.2.2.6.1 Lambda DNA preparations

λ DNA was prepared using methods as described by Aber *et al.* (1983). The λ lysate was divided into 250ml Sorval bottles. λ bacteriophages were precipitated in the lysate by adding NaCl to 6% and PEG6000 to 11%. The bottles were inverted repeatedly to ensure the solids dissolved. The bottles were stored on ice for 1 hour and then spun at 10000 rpm for 10 minutes to pellet the phage. The phage/debris pellet was resuspended in 10ml of phage suspension buffer and the liquid transferred to a Falcon tube. 0.25 vols of 20% PEG, 2.5M NaCl solution were added to the phage and the Falcon tube stored on ice for 5 minutes. After another spin, the pellet was resuspended in 0.5ml of phage suspension buffer and DNase 1 and RNase A added. The tube was incubated at 37°C for 15 minutes and a chloroform extraction carried out. The supernatant was then adjusted to 0.2% SDS, 20mM EDTA and mixed carefully. The SDS breaks up the phage protein capsid and the EDTA chelates divalent cations necessary for the function of nucleases. After heating the tube at 65°C for 5 minutes a phenol chloroform extraction was carried out followed by a chloroform extraction. The supernatant was then precipitated using 3M NaAc and isopropanol and the lambda DNA pellet washed with 70% ethanol, dried and resuspended in 100µl of TE.

2.2.2.6.2 Fungal genomic DNA preparations

Fungal genomic DNA for use in Southern Blots was extracted by two methods.

METHOD 1

The first method used was that of of Borges *et al.* (1990). 0.5-2g of freeze-dried mycelium were ground up and the powder divided between two Sorvall tubes containing 15ml of cold spermidine-SDS buffer. The tubes were shaken and the mixture extracted twice with phenol followed by a chloroform extraction. The DNA was ethanol precipitated, washed with 70% ethanol, dried and resuspended in 1ml TE.

METHOD 2

The second method used was one adapted from a protocol by Xu *et al* (1994). 200mg of freeze-dried mycelium were divided into 10 tubes and coarsely ground. In a fume cupboard 600µl of extraction buffer were added to each tube. The tubes were left for 30 minutes and then an equal volume of chloroform-isoamyl alcohol 24:1 was added to each tube. The tubes were mixed vigorously for 5 minutes then spun at 13000 rpm for 5 minutes. After this, the DNA was precipitated using 550µl of cold isopropanol and spun again for 1 minute. The pellets were dried by inverting the tubes on absorbant paper and then dissolved in 100µl of TE. The tube contents were then pooled into two groups. Three extractions were carried out using phenol, then phenol-chloroform and finally chloroform-isoamyl alcohol 24:1 to remove residual nucleolytic activity. The DNA was then precipitated using 90µl of 7.5M NH_4Acetate and 900µl of cold 100% ethanol. The DNA pellets were washed with 70% ethanol, dried and then dissolved in an appropriate volume of TE, depending on the size of pellet obtained. Method 2 was found to remove the carbohydrate from the DNA more effectively than the first method of DNA extraction.

2.2.2.6.3 Plasmid DNA preparations.

METHOD 1 Alkaline Lysis

DNA for performing automated sequencing was prepared by this method. It is based on the alkaline lysis method of Birnboim and Dolly (1979). Using the host strain XL-1 Blue, single colonies containing a plasmid were picked and grown up overnight shaking at 37°C in 10ml of L-Amp. 4ml of overnight culture in total were spun down in an eppendorf tube. The bacterial pellet was resuspended in 200µl of GTE buffer. 300µl of freshly prepared 0.2M NaOH/ 1% SDS were added, the tube mixed by inversion and incubated on ice for 5 minutes.The solution was neutralised by adding 300ml of 3M potassium acetate, pH4.8, and the tube again mixed by inversion. The tube was incubated on ice for a further 5 minutes. Cellular debris were removed by spinning the tube for 10 minutes at room temperature, then transferring the supernatant to a clean tube. The preparation was RNased at 37°C for 20 minutes and the supernatant extracted using 400µl of chloroform. This extraction step was repeated. The DNA was precipitated using an equal volume of 100% isopropanol and the preparation immediately centrifuged for 10 minutes at room temperature. The DNA pellet was washed using 500µl of 70 % ethanol and dried under vacuum for 3 minutes. The pellet was dissolved in 32µl of deionized water and the plasmid DNA precipitated by adding 8µl of 4M NaCl first followed by 40µl of autoclaved 13% PEG8000. The preparation was mixed thoroughly and incubated on ice for 20 minutes. The plasmid DNA was pelleted by centrifugation for 15 minutes at 4°C. The pellet was rinsed with 500µl of 70% ethanol, dried for 3 minutes under vacuum and resuspended in 20µl of deionized water.

METHOD 2 Boiling method

This method uses the protocol of Holmes and Quigley (1981). 4ml of overnight culture, XL-1 Blue again was the host strain, were spun down in an Eppendorf tube. The pellet was resuspended in 200µl of STET and 20µl of lyzozyme (10mg/ml). The suspension was boiled for 40 seconds and centrifuged for 10 minutes. The flocculent pellet obtained was removed with a Gilson tip. The supernatant was then precipitated using 3M NaAc and 200µl of isopropanol. After placing the tube at -70°C for 10 minutes, the pellet was rinsed with ether and dried before resuspension in 50µl of TE.

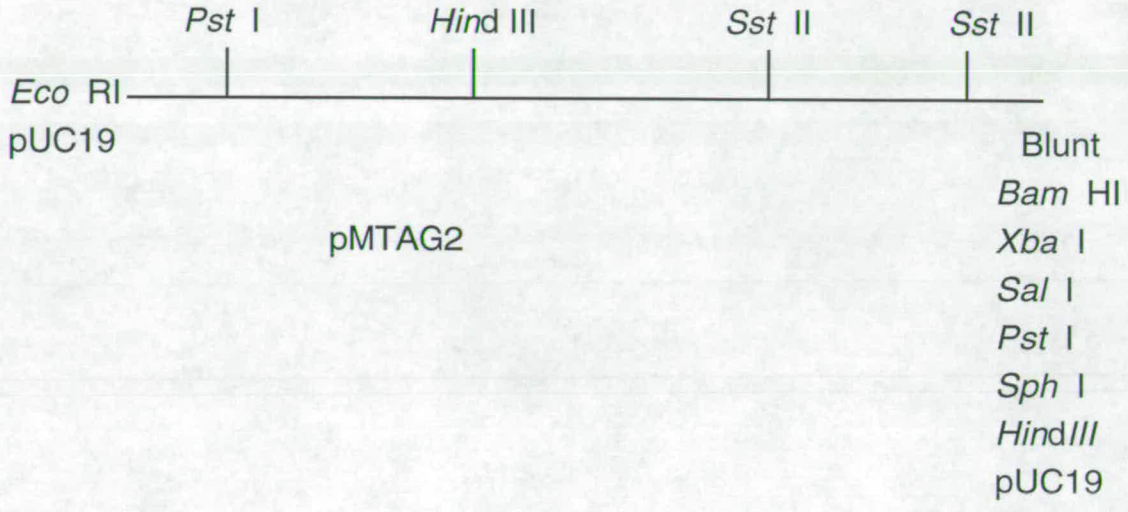
2.2.2.7 Southern Blots

Southern Blots were carried out using the protocol of Southern (1975). The gel from which DNA was to be transferred was soaked in 0.2M HCl for 20 minutes, which aids transfer of fragments above 15kb as the HCl depurinates the DNA. The gel was then soaked in denaturing solution for 30 minutes followed by a soak in neutralising solution for 1 hour. 20X SSC was used as the transfer buffer, this solution providing a high ionic strength to ensure efficient binding of small fragments to the nitrocellulose filter. Two pieces of blotting paper were soaked in 20X SSC and put on a glass sheet on a tray containing 20X SSC so that the ends of the blotting paper dipped into the liquid and acted as wicks. The gel was then placed on the blotting paper and the nitocellulose membrane soaked in 2X SSC carefully placed on top. This was covered with a sheet of blotting paper also soaked in 2X SSC and several dry sheets. A pad of paper towels held down with a weight was placed on top of the blotting paper. The gel was left to transfer overnight, then the membrane was rinsed in 2X SSC and left to bake for 2 hours at 80°C.

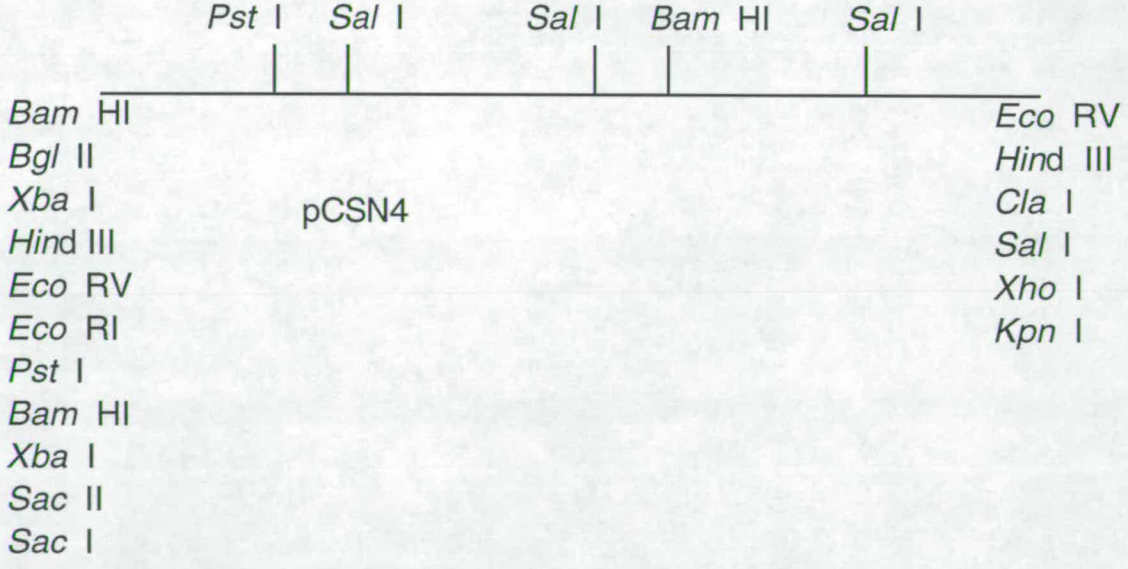
2.2.2.8 Preparation of probes

Subclones of the mtA and mt**a** *N.crassa* mating type genes were cloned into pMTAG2 and pCSN4 respectively. The mating type specific fragments are shown below.

A mating type specific: 1.2kb *Eco* RI/*Bam* HI fragment



a mating type specific: 1.9kb *Eco* RV fragment



Digestion of pMTAG2 with Eco R1/Bam H1 gives a fragment of 1.2kb which is A mating type specific. Digestion of pCSN4 with Eco RV gives a fragment of 1.9kb which

is **a** mating type specific. The bands were excised from agarose as described in 2.2.2.4 and the DNA pellet resuspended in 10µl of TE.

2.2.2.9 Labelling the probes

Labelling of the probes was carried out using the Pharmacia Biotech Oligolabelling kit. 50ng-1µg of DNA was first denatured by heating at 100° and placed on ice. 25µl of distilled water were added to the DNA along with 10µl of reagent mix. The reagent mix contains hexadeoxyribonucleotides of random sequence which anneal to random sites on the DNA and serve as primers for DNA synthesis by the Klenow fragment of *E.coli* DNA polymerase I. 4µl of ^{32}P dCTP probe were added so labelled nucleotides were incorporated during the DNA synthesis. The 1µl of Klenow was added last to make a total of 50µl. The mixture was left overnight at room temperature to allow the label to incorporate into the DNA.

Unincorporated radioactivity was removed from the probe by passing the 50µl down a Stratagene Nuctrap column. The probe was then denatured by boiling before being added to the hybridisation mixture.

2.2.2.10 Hybridisations

Hybridisations were carried out in Techne hybridisation cylinders placed in Techne ovens. Hybridisations were carried out overnight. Lambda and plasmid DNA blots were generally hybridised at 62°C and washed at 57°C. Genomic DNA blots were usually hybridised at 58°C and washed a few degrees lower, depending on background levels of radioactivity present on the filter. Hybridisations were carried out in a hybridisation mix as described in the materials section. 10% dextran sulfate was added to accelerate the rate of hybridisation and 5X Denhart's reagent added to reduce background when carrying out hybridisations with genomic blots. After hybridising overnight the blots were washed three times at 30 minute intervals using the wash solution described in the Materials section.

2.2.2.11 Colony screening

Many vectors carry a segment of *E.coli* DNA that contains the regulatory sequences and amino-terminal coding information for the ß-galactosidase gene (*lac* Z). In the coding region there is a polylinker site into which a desired insert can be cloned. A vector like this is used in a host cell which codes for the carboxy-terminal portion of *lac* Z. The host encoded fragment and the vector encoded fragment can associate to form an enzymatically active protein. The Lac^+ bacteria resulting from this complementation are blue in the presence of IPTG and X-gal. Those vectors into which the desired insert has been cloned are identified as white colonies, they are unable to complement with the host fragment due to the presence of the insert in the polylinker site.

For those vectors not carrying a segment of *lac* Z, a colony lift can be taken using nitrocellolose and the colonies denatured and neutralised (Benton and Davis, 1977). The filter can then be probed using hybridisation techniques described previously so colonies with the desired insert are recognisable on an autoradiograph.

2.2.2.12 Sequencing

MANUAL SEQUENCING

Sequencing was carried out using the Sequenase Version 2.0 DNA sequencing kit supplied by United States Biochemicals. The method this kit uses follows the protocol of Sanger *et al* (1977). Approximately 3µg of DNA are required for each reaction. DNA templates were double stranded so the DNA first had to be denatured in a total of 100µl using 0.2M NaOH and 0.2mM EDTA. The DNA was left to incubate at 37°C for 30 minutes and then ethanol precipitated. The DNA was resuspended in 6µl of TE. The primer being used was then annealed to the DNA template. 1µl of primer, approximately 1pmol in concentration, was added to the DNA along with 2µl of Reaction buffer as supplied and 1µl of DMSO. This mixture was left to incubate at 37°C for 30 minutes. The labelling reaction was carried out next where 1µl of 0.1M DTT, 2µl of Labelling mix diluted 1:5 with distilled water, 0.5µl of ^{35}S dATP and 2µl of Sequenase Version 2.0 diluted 1:8 in Sequenase dilution buffer were added to the annealed template-primer. The reaction was left to incubate for 2-5 minutes at room temperature after which 3.5µl were removed and added to an Eppendorf containing 2.5µl of ddGTP. The same procedure was carried out again using tubes containing ddATP, ddTTP and ddCTP. The Sequenase Version 2.0 synthesises a DNA strand using the primer as the initiation site. The

synthesis reaction is terminated by the incorporation of the ddNTPs so a population of chains of different lengths is generated. For example in the case of ddTTP DNA synthesis will halt at each thymine generating a population of chains of different lengths. The four separate reactions with each ddNTP will give the complete information needed for the DNA fragment being sequenced. The termination reactions were incubated at 37° for 5 minutes and 4µl of Stop solution were added. When the gel on which the samples are to be run was ready to be loaded, the samples were heated at 75-80°C for 2 minutes and 3µl loaded.

AUTOMATED SEQUENCING

Automated Sequencing was carried out using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit supplied by Perkin Elmer. In a 0.5ml tube the following reagents were added: 8µl of Terminator Ready Reaction Mix as supplied, 1.5-2.5µl of template, 3.2pmol of primer and distilled water to 20µl. The Terminator Ready Reaction Mix contains A, C, G andT-Dye Terminators, dITP, dATP, dCTP and dTTP, Tris-HCl pH9, $MgCl_2$ thermal stable pyrophosphatase and AmpliTaq DNA polymerase, FS. The mixture was overlayed with a drop of mineral oil. A Hybaid OmniGene programmable PCR machine was programmed for thermal cycling following the instructions in the User's Manual. Extension products were purified as follows: the 20µl from the reaction tube was transferred to a tube containing 2µl of 3M Sodium acetate, pH 4.6, and 50µl of 95% ethanol. This tube was vortexed, placed on ice for 10 minutes and centrifuged at maximim speed for 15-30 minutes. After removing the supernatant the pellet was rinsed with 250µl of 70% ethanol and dried in a vacuum centrifuge. The pellet was resuspended in loading buffer containing 5 parts deionized formamide to 1 part 25mM EDTA, pH8 + 50mg/ml Blue dextran. The sample was vortexed and spun, heated at 90° for 2 minutes and placed on ice before loading. The sample was run on an ABI PRISM 377 DNA Sequencer.

2.2.2.13 Acrylamide gels and equipment used for manual sequencing

Stock Acrylamide was bought pre-made from Scotlab.The Easigel mix consists of 6% acrylamide/0.3% w/v bis acrylamide, 7M Urea and 1X TBE. Biorad sequi-gen sequencing cells were used to run the sequencing samples. The sequencing gels were run using 1X TBE buffer.

2.2.2.14 Polymerase Chain Reaction

The Polymerase Chain Reaction was used to amplify mtA-1 of *Sordaria equina*.. The following table describes the concentration and volume of the components used. The *taq* DNA polymerase (5U/µl) was supplied by GIBCOBRL. The components were added on ice.

	µl Control	µl Control	µl 1/10 dilution of template	µl 1/100 dilution of template
Template (100ng/µl)	0	1	1	1
10X PCR buffer	10	10	10	10
dNTP (100mM)	0.7	0.7	0.7	0.7
Upstream primer (20ng/µl)	10	10	10	10
Downstream primer (20ng/µl)	10	-	10	10
Taq DNA polymerase	0.5	0.5	0.5	0.5
Mg^{2+} (50mM)	3	3	3	3
Distilled water to 100 µl	65.8	74.8	64.8	64.8

The mixture in each tube was then covered with 75µl of mineral oil to prevent evaporation. The following specifications were used during 21 cycles of PCR amplification. Reactions were carried out using a Hybaid OmniGene PCR machine.

Denaturation	Annealing	Polymerisation
95° 5mins	55° 1min	70°C 2mins
93° 3mins	55° 1min	70°C 2mins
93° 3mins	55° 1min	70°C 2mins
cycles 4-20		
21		70°C 5mins

The amplification product was checked by running a small quantity of the PCR reaction against molecular weight standards on an agarose gel.

*Taq*I polymerase adds a 3' terminal adenine to both strands. The pGEM-T Vector Systems (Promega) takes advantage of this and following the manufacturer's directions, the *S. equina* mtA-1 PCR product was "TA" cloned into the pGEM-T Vector.

2.2.3 Fungal Manipulations

2.2.3.1 Making spheroplasts from conidia

30ml of sterile distilled water were added to a "swirl flask" of *N. crassa.* The water was swirled gently to harvest the conidia. The suspension was poured through two layers of cheesecloth into a sterile 50ml tube. The tube was spun at 3000 rpm for 10 minutes and the supernatant removed. The pellet was resuspended in 30ml of sterile distilled water. Once more the conidia were spun down at 3000 rpm for 10 minutes and the pellet suspended in 5ml of sterile distilled water. The conidia were counted on a haemocytometer and 2X 10^9 conidia innoculated into 150ml Vogels N medium + 1.5% sucrose + adenine. The conidia were allowed to germinate overnight at 200 rpm while shaking at 15°C. The next day the conidia were spun down at 3000 rpm for 5 minutes and the supernatant removed. The conidia were washed by resuspending in 50ml of distilled sterile water and the centrifugation repeated. This time the pellet was resuspended in 10ml 1M Sorbitol and 4ml of filter sterilised Novozyme (5mg/ml in 1M Sorbitol) were added to digest the cell walls of the conidia. The Sorbitol prevents the spheroplasts bursting. The conidia with the Novozyme were incubated with shaking at 100 rpm, 30°C for 1 hour. 50ml of 1M Sorbitol were added after the 1 hour period and the spheroplasts centrifuged and resuspended in 50ml of 1M Sorbitol three times to wash the spheroplasts free of Novozyme. The spheroplasts were suspended in a final volume of 2ml of 1M Sorbitol. The spheroplasts were then ready for electroporation. The spheroplasts were also diluted serially (10^{-2}-10^{-5}) and added to overlay medium without Sorbitol (this checks that the spheroplasts have no cell wall as they should not grow) and to overlay medium with Sorbitol (this checks the viability of the spheroplasts). The overlay was then poured onto plates containing non-selective underlay medium as described in section 2.1.3.

2.2.3.2 Electroporation of Spheroplasts.

40µl of the spheroplast suspension was gently mixed with 0.1µg of plasmid pBARGRG1JN and left on ice for 5 minutes. The BIORAD Gene Pulser apparatus was set to 1.5kV and 25µF with the pulse controller at 200Ω. The spheroplast/DNA mixture

was then transferred to a 0.2cm electroporation cuvette and the cuvette placed in the safety slide. The slide was pushed into the chamber and a pulse applied. The spheroplast/DNA mixture was immediately transferred into 1ml of cell wall regeneration mixture (see section 2.1.3) and the mixture allowed to shake at 200rpm for 3 hours at 30°C. After this period of time the mixture was directly plated onto selective medium and incubated at 28°C. Any colonies that grow should be transformants which are then transferred onto selective slants.

2.2.4 **Computer Analysis**

2.2.4.1 GCG programmes.

Programmes from the Wisconsin package GCG version 9 were used to analyse DNA and protein sequences and compare them to homologous sequences present in the DNA and protein databases.

2.2.4.2 Phylogenetic analysis.

The CLUSTALW programme (Thompson *et al*, 1994) was used to carry out the multiple alignment of the mtA-1 DNA and protein sequences. The DNA sequence data contains more information than the protein sequence so the phylogeny shown in chapter 6 was based on the DNA sequence data. However the phylogeny based on the protein sequence data (not shown) had the same topology. The phylogenetic analysis was carried out under the guidance of Dr. Frank Wright (Biomathematics and Statistics Scotland, based at the Scottish Crop Research Institute in Dundee).

Huelsenbeck (1995) reviewed the three classes of method for constructing phylogenetic trees, maximum parsimony, distance methods and maximum likelihood, and concluded that all three methods were successful when the true phylogenetic tree contained no long branches and when the rate of evolution did not vary among sites i.e all the positions in the sequence evolve at the same rate. When these two conditions occur then the best method to use is maximum likelihood, if the model of evolution assumed is correct. The model used should allow for different rates of transitions from transvertions and for a variable rate of evolution at different sites along the sequence.

The maximum likelihood method was therefore used to analyse the DNA alignment. The DNA Maximum Likelihood program (DNAML) was available on the PHYLIP 3.6

package. A paper discussing the methodology and the DNAML program has also appeared in the literature (Felsenstein and Churchill, 1996).

The DNAML programme was used to estimate the relative rates of transitions and transvertions. This is summarised as "the expected transition/transvertion ratio" and a value of 1.5 was obtained for the dataset. The three codon positions and the intron are likely to evolve at different rates (although the third position and the intron may have similar rates). The relative rates of these four categories were estimated by splitting the alignment into four subalignments containing only one category of sites in each, then calculating the average pairwise genetic distance (using the DNADIST program in PHYLIP 3.5) for each category. The ratios of these average distances were calculated relative to the second codon position. This produced relative rates of 1.61 : 1 : 2.33 : 2.31 for the first, second and third codon positions and the intron, respectively.

Most of the among site variation may be due to between category rate variation. However it is likely that there is additional rate variation. One can imagine for example that a protein coding region of a sequence may be more conserved than an intron. DNAML can fit different rates to different "regions". With the number of regions chosen the relative rates of these regions and the probability that a particular site belongs to that region can be calculated. Two classes of regions were chosen, with relative rates 5:1 and probabilities of 0.5 and 5.0. The relative rate values (for categories and regions) plus the optimised transition/transvertion ratio were used to do a final run of the DNAML program with a careful search for the most likely tree.

To test the robustness of the tree, a statistical procedure called nonparametric bootstrapping was carried out. The bootstrap method places a level of support for each clade in the tree and values greater than 70% are generally thought to provide good support for the clade (see Swofford *et al* , 1996 for further discussion). The bootstrap analysis was carried out using the SEQBOOT and DNAML programs from PHYLIP version 3.6.

Chapter 3

The Sub-Cloning and Sequencing of *Sordaria sclerogenia* MtA-1.

3.1 Introduction

S. sclerogenia is a heterothallic species where mtA individuals mate with mta individuals. A lambda EMBL4 clone from a *S. sclerogenia* mtA genomic library probed positive with the *N. crassa* mtA probe (see Materials and Methods). L.Bisoni carried out the first stages of sub-cloning the mtA-1 gene from the lambda clone. MtA-1 was then sequenced as was the common region and part of the variable region.

3.2 Results

3.2.1 The Sub-Cloning of MtA-1 of *S. sclerogenia.*

Figure 3.2.1 a) shows how an *Eco* RI/*Xho* I fragment which hybridised with the *N. crassa* mtA probe was subcloned into a pBluescript KS+ plasmid digested with *Eco* RI and *Xho* I. Further subcloning to enable mtA-1 to be sequenced was carried out by L. Bisoni and myself taking advantage of the *Pst* I, *Hind* III and *Xho* I sites as shown in figure 3.2.1 b). Fragments to be sub-cloned were cut from low-melting point agarose gels, phenol extracted and ligated into appropriately digested pBluescript KS+. The plasmid was then transformed into XL-1 Blue competent cells and plasmids with inserts selected using blue-white selection.

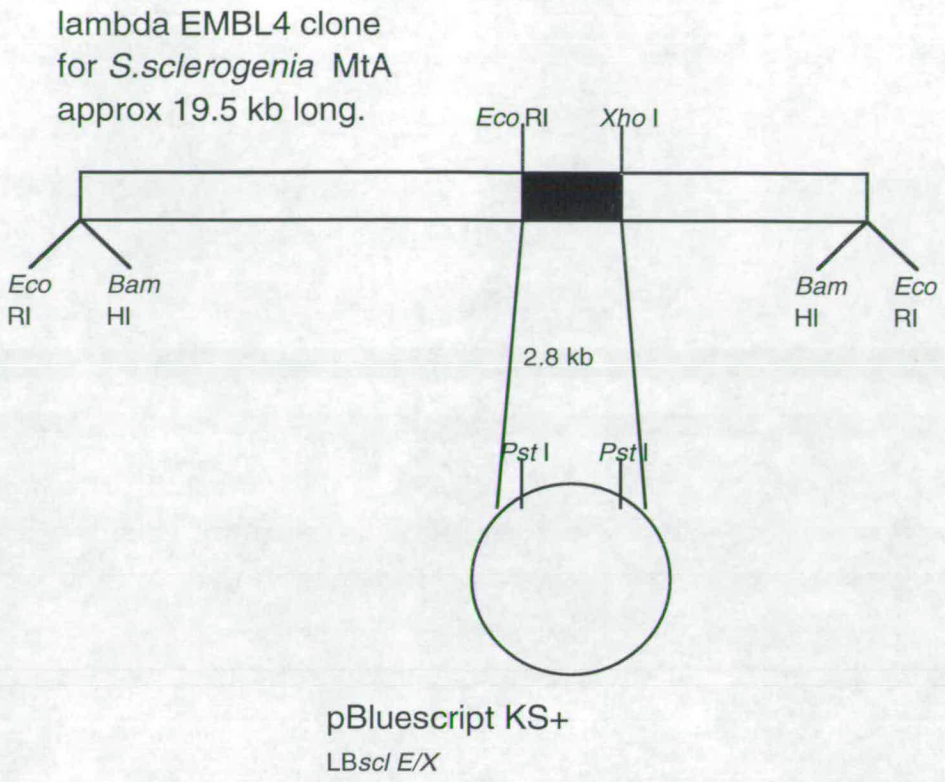


FIGURE 3.2.1 a) Initial work carried out by L. Bisoni. The 2.8 kb *Eco* RI/*Xho* I fragment which hybridised with *N. crassa* MtA probe is shown as a black rectangle. The subcloning of this fragment from lambda is represented schematically here.

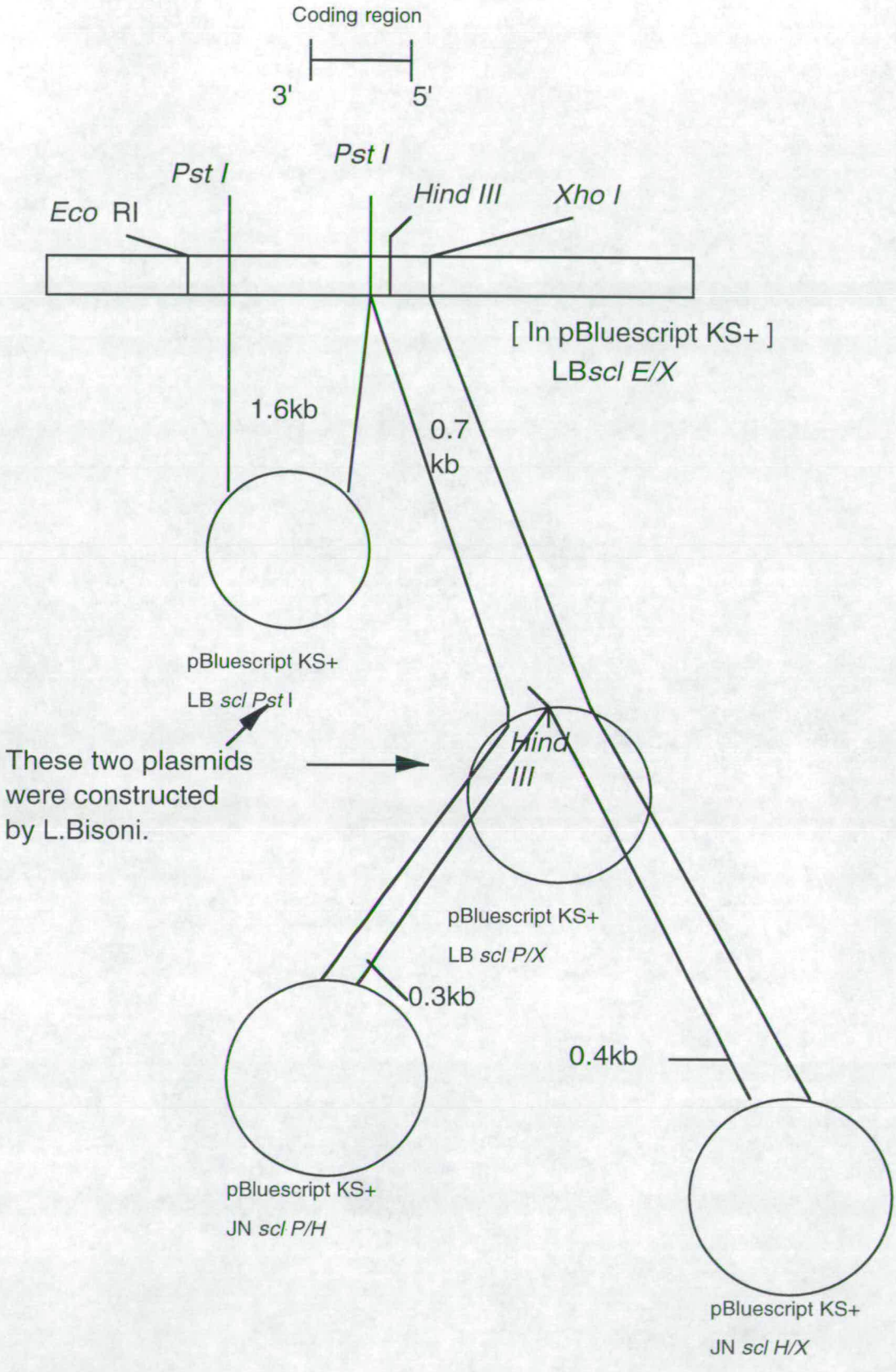


FIGURE 3.2.1 b) Schematic diagram of the sub-cloning of MtA-1 of *S. sclerogenia.* The position of the MtA-1 coding region is shown.

3.2.2 Sequencing of *S. sclerogenia* MtA-1, Common Region and Variable Region.

Double-strand sequencing was performed on DNA prepared by the boiling method. Automated sequencing was performed on DNA prepared by alkaline lysis. Standard pBluescript KS+ primers were used aswell as primers ordered from Oswel which annealed to the MtA sequence. Figure 3.2.2 a) shows the DNA sequence of MtA-1, the common region and part of the variable region of *S. sclerogenia*. Figure 3.2.2 b) shows the deduced amino acid sequence of mtA-1.

1 tttttcccaa cattagcctc gcaatcgaat ctcgcctgca cttcctcacg

51 tgttgaaccc tcaccattca agaactcgaa cgccagaaac gcgatgtcag

101 gcgtcgatca aatcgtcaaa acgttcgcca atctcggtga gggcgatcgt

Hind III

151 gaagcggcca tgaa*agctt*t cttagcgatg atgcccgtga gcaacgaaac

201 tgtcgccgaa cctgtccgca aagcccctgc cgcgaagaag aaggtcaacg

251 gcttcatggg tttcagatgt aagtcagatc tgagtcaatc tcgtggacag

301 tccatactaa ttgtttttc ttcagcgtac tattccccgc tcttctctta

351 ccttccacaa aagatgaggt cgcctttcat gaccattctc tggcagtacg

401 atccctacca caccgaatgg gatttcatgt gttcggttta ttcctcaatc

Pst I

451 cgcaccgacc tggaggagga gaaggttaca *ctgcag*ctct ggattcacta

501 tgctatcggc cagatgggat tgattgaccg cgaccactac atggcatcgt

551 ttggctggcg cctcggtcag actcgcaacg gcactaccga ccttttcgc

601 actgcgattc cgatggttag gcgcaagctt cagcccatga acggcctttg

651 cctgctcatt aagtgtctcc agagcggatt gaacaagcat cttaccaatc

701 ctcatcctgt tattgccaag ctgcaagatc ctagcttcga catgatttgg

751 atcaacaagc cttctcacca tcagcaggga cacaccgatc aagctgacaa

801 ttctgaactc agaatgccgt cgatcttccc tagcaatcac gcagtcgctg

ccttcctc tctcacattg

851 cggaggtaga tggcatcgcc aaccttcctc tctcacattg gactcagcag

901 ggggatttcg gcaaggagcc tggattctcg acccagtttg ataccatgtt

951 ggattcactt cttgagaatg gaaacgacac cagcaatcat cactacaaca

1001 tggctctggc tatggatctt cccatgatag gttagtggaa gacgaggcac

1051 catgtcgctt acgttcaacc cgtgtgctga ccatttggca ggatttaatg

1101 gaggagcata aaagcacggc gcagttacgg tttcttttcc tttgtcacat

1151 ttgggtttcg tggttcaggc atacaaagcg agggcgaaaa ggggtctagt

1201 tgggtttctt tgtgcattca cttggggcaa atcataggac ttcagaatcg

1251 aacgttgtgg aataggcaat taaacggcag acaggtagct agctgcctaa

```
1301  ctagatggca agagcaaatg aAATCAATga cagtcaacag cgtccactat

1351  ctggtttagg atggtcttca ctagtgaatc gatatgaacg tactgtgtaa
      gcgtctgcgg tatcgttg
1401  gcgtctgcgg tatcgttgcc gaatttgaca tgtcgtcgag ataaagctgt
                                    Nsi I
1451  ggtccgccac tgacgccaat gcttatgcat gcagccttac gctagtgcga

1501  atccgtcgct gggcagtacc tgcctcgtgc atgacaaaat taagggaacc

1551  accaaaccgg tttcctccct gttgatgatg agatctgccg tgagctcgta

1601  gaggaaggct gggctgctct ctctcaagaa cggataaatt attggataga

1651  tatgatccct caaatcctcc aagatggtat agatctaggg ggagcattaa

1701  cagctcacta aagcgatgaa aacgaagtct tttggtaaag aaacggccga

1751  gttctggctc atcaaagtct ggggtggcct gtgcacggtt cccatgattt

1801  tgtcttgaac gagcctgcgt atgtacgcta gtgtagaagc cgtgcccctc

1851  atttatggat cttgttcctt gacctgcagg
```

FIGURE 3.2.2 a) The sequence of *S. sclerogenia* mtA-1, the common region and part of the variable region. The mtA-1 coding region is shown underlined. The putative mtA-1 intron, nucleotides 269-325, is shown in plain text. The common region, nucleotides 1423-1480, is shown underlined in italics The variable region follows directly on from the common region in plain text. Primers which were ordered (Oswel) to complete the sequencing are shown in italics above the corresponding sequence. Standard pBluescript KS+ primers were used elsewhere. Restriction sites are shown also in italics. The putative polyadenylation signal is underlined in upper case.

```
  1  MSGVDQIVKT FANLGEGDRE AAMKAFLAMM PVSNETVAEP VRKAPAAKKK

 51  VNGFMGFRSY YSPLFSYLPQ KMRSPFMTIL WQYDPYHTEW DFMCSVYSSI

101  RTDLEEEKVT LQLWIHYAIG QMGLIDRDHY MASFGWRLGQ TRNGTTDLFR

151  TAIPMVRRKL QPMNGLCLLI KCLQSGLNKH LTNPHPVIAK LQDPSFDMIW

201  INKPSHHQQG HTDQADNSEL RMPSIFPSNH AVAAEVDGIA NLPLSHWTQQ

251  GDFGKEPGFS TQFDTMLDSL LENGNDTSNH HYNMALAMDL PMIG*
```

FIGURE 3.2.2 b) The deduced amino acid sequence of *S.sclerogenia* mtA-1. The region proposed to be a DNA binding domain with similarities to the α domain of *S. cerevisiae* MATα1 and *N. crassa* mtA-1 is shown underlined.

3.3 Discussion

The *Sordaria sclerogenia* mtA-1 gene, using the definition of mtA-1 by Glass *et al* (1990a) comprises of two exons separated by an intron 57bp long which is found 176bp downstream of the first ORF ATG start site. The position of the intron is conserved in relation to the position of *N. crassa* mtA-1 intron (see chapter 6). The intron has a 5' splice site GTAAGT and 3' splice site CAG. A putitive polyadenylation sequence is located at nucleotide 1322, AATCAAT. Polyadenylation signals are usually 11-30bp upstream of the polyadenylation site and have a consensus sequence of AAUAAA. If the putative intron were spliced out, the first and second exons would code for a translational product encoding a 294 amino acid polypeptide. Comparison of the *S. sclerogenia* mtA-1 DNA and polypeptide sequences to *Neurospora* mtA-1 sequences is shown in chapter 6.

Recent work by Saupe *et al* (1996) has identified the presence of a second intron in the *N. crassa* mtA-1 ORF by 3' RACE experiments. This intron was found to be 60bp long with a 5' splice sequence GTTAGT and 3' splice sequence CAG. When this intron is spliced from mtA-1, five amino acids are added to the 288 amino acid mtA-1 ORF as previously defined by Glass *et al* (1990a). Using a primer located in the second intron the mtA-1 cDNA encoding the 288 amino acid polypeptide was amplified by RT-PCR

(Glass *et al*, 1990a) suggesting that the second intron is not removed in all of the mtA-1 transcripts. Saupe *et al* (1996) proposed that this reflects the multi-functional nature of mtA-1 in mating and in post-fertilization functions. In *S. sclerogenia* the 5' splice sequence of this second intron is conserved (GTTAGT, nucleotides 1031-1036 figure 3.2.2 a)) as is the 3' splice sequence (CAG, nucleotides 1089-1091 figure 3.2.2 a)). The second intron is 61bp long. Five amino acids can be added to the *S. sclerogenia* mtA-1 ORF when this intron is spliced out. The stop codon is located at nucleotides 1109-111, TAA. In this chapter *S. sclerogenia* mtA-1 is defined as a 293 amino acid ORF as opposed to a 298 amino acid ORF because it is not known what percentage of the transcripts have this second intron spliced out.

Chapter 4

The Sub-Cloning, Sequencing and Expression of *Sordaria equina* MtA-1.

4.1 Introduction

S. equina is a homothallic species containing mtA only. A λ EMBL4 clone from a *S. equina* genomic library probed positive with the *N. crassa* mtA probe. S. Liddle subcloned a 1kb *Xho* I fragment from the lambda clone which probed positive with the mtA probe. MtA-1 was further subcloned and sequenced as was the common region and part of the variable region. MtA-1 was cloned into an expression vector, pBARGRG1, and transformed into sterile mt**a** *N. crassa* spheroplasts to observe whether the *S. equina* mtA-1 could restore mating type function.

4.2 Results

4.2.1 The Sub-Cloning of MtA-1 of *S. equina.*

Figure 4.2.1 shows how the mtA-1 gene of *S. equina* was sub-cloned into plasmid vectors for sequencing. Originally the 1kb *Xho* I fragment sub-cloned from the lambda clone by S. Liddle was further sub-cloned by myself taking advantage of the *Sal* I site which cut the *Xho* I fragment and made sequencing the fragment easier. The sequencing data obtained from the *Xho* I fragment showed that approximately 360bp of the downstream portion of mtA-1 was not present on the fragment. A 0.28kb *Xho* I fragment and a 3kb *Xho* I/ *Bam* HI fragment were sub-cloned to obtain the missing sequence data. Fragments were sub-cloned as described in figure 4.2.1.

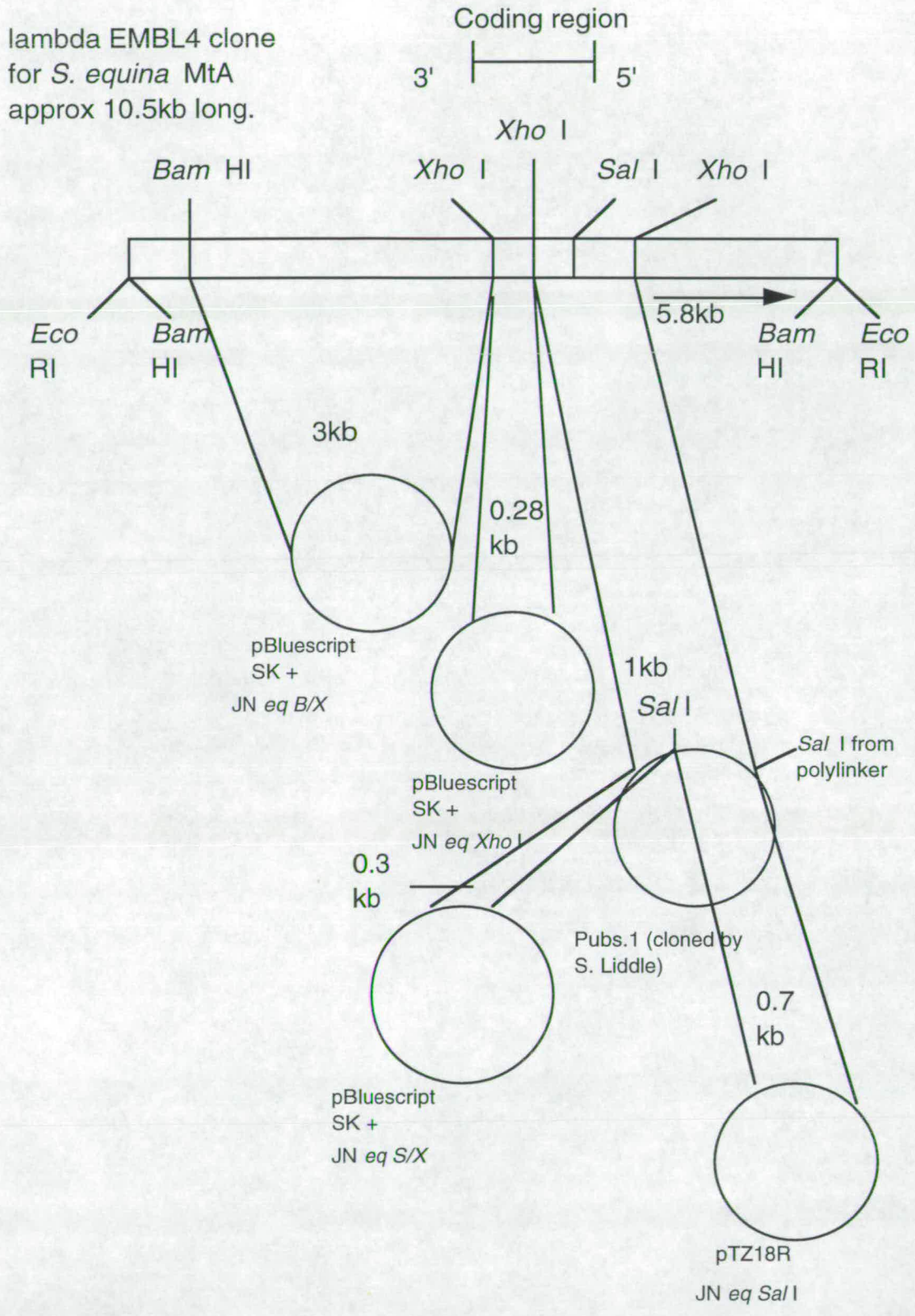


FIGURE 4.2.1 Schematic diagram of the sub-cloning of MtA-1 of *S. equina*. The position of the MtA-1 coding region is shown.

4.2.2 Sequencing of *S.equina* mtA-1, Common Region and Variable Region.

Double-strand sequencing was performed using DNA prepared by the boiling method. Standard pBluescript SK+ and pTZ18R primers were used aswell as primers ordered from Oswel which annealed to the mtA sequence. Figure 4.2.2a) shows the sequence of MtA-1, the common region and part of the variable region of *S. equina.* Figure 4.2.2b) shows the deduced amino acid sequence of mtA-1.

```
   1  acatacaatg ctagtttcta gcaaggaatg aaatcataat tcgatcaaat

  51  tgctttcact tttacctgtc aaattcacaa ggctgcacgt cttatgtaat
                                       a agcagcaaac ctacccg
 101  ccaaaagccc ttgtgaagtt gcgcccccaa agcagcaaac ctacccgcct

 151  tctccctccc ccgcctcccc gtccctcggt ccgtaagtga atggaaagga

 201  gacagaaaac gcgcccaccC AAATtaacag tcaaccccat gtctcctatt

 251  taggaaagcc gtggtcatct ttcccgcctt cacccaaact tcccaccact

 301  tttttcccaa catcagcctc gtaatcgaaa tctcgtctgc acttcctcac

 351  gtgttgaact catcattcaa gaactcaaac gccagaaacg cgatgtcagg

 401  cgtcgatcaa atcgtcaaaa agttcgccaa tctcggtgag ggtgatcgtg

 451  aagcggccat gaaagctttc ttagcgatga tgcccgtgag caacgaacct

 501  gtcgctgaac ctgtccgcaa agcccccacc gcaaagaaga aggtcaacgg
                                                      Sal I
 551  cttcatgggt ttcagatgta agtcagatct gagtcaatct tgtcgacagt

 601  ctatgctaat tgtttttcct tcagcgaact attccccgct cttctcttac

 651  cttccgcaaa agatgaggtc gcctttcatg accattctct ggcagtacga

 701  cccctaccac aacgaatggg atttcatgtg ttcggtgtat tcttcaatcc

 751  gcaccgacct ggaggagcag aatgttacac tgcagctctg gattcactat

 801  gctatcggcc agatgggatt gattgaccgc gaccactaca tggcatcgtt

 851  tggctggcgc ctcggtcaga ctcgcaacgg cactaccgac ctttttcgca

 901  ctgcgattcc gatggttagg cgcaaccttc agcccatgaa cggcctttgc
                      Xho I
 951  ctgctcatta agtgtctcga gagcggattg agcaagcatc ttaccaatcc

1001  ccatcctatt atcgccaagc tgcaagatcc tagcttcgac atgatctgga

1051  tcaacaagcc tcctcaccat cagcagggac acaccgacca agctgacaat

1101  tccgaactcg gaatgccgtc gctcttccct ggaaatcatg cagtcgctgc

1151  ggaggtagat ggcatcgcca accttcctct ctcacatcgg actcagcagg

1201  gagatttcgg caccgagcct ggattctcga ctcagtttga taccatgttg
                Xho I
1251  gattcaattc tcgagaatgg aaaccatccc agcaatagtc actacaacat

1301  gtctctggct atggatcttc ctatgacggg ttagtggaag acgaggcacc

1351  atctcgttta cgttcaactc gtgtgctttc acttggggca aatcatggga
```

```
1401  cttcagaatc gaacggtgtg gaatggacaa tcaaacggca gacaggtagc
                                                        gcgt
1451  tacctaacta gatggcaaga gcaAATAAAA tcaatgacag tcaacagcgt
      ccactatctg gttt
1501  ccactatctg gtttaggatg gtcttcacta gtgaatcgat atgaacgtac

1551  tgcttaagcg actgcgtatc gttgccggat ttgacatgtc gtcgagataa
                                        Nsi I
1601  agctgtggtc cgccactgac gccaacgctt atgcatacaa cctcacgcta

1651  gcgtaaatcc gtcgctgggc agtacctgcg tatgtacact agcgtagaag

1701  ctgtgcccct catttatgga tctgctaatg tttcttgacc tgcagtttct
      ggagatccgc atcagtcc
1751  ggagatccgc atcagtccat tggagaagcc gcataccgtc ttcatttcac

1801  tagtacttct acacagtcga gcccgttgca aaatcgaagt gtacttaaca

1851  gtcagagtga ctaaatgagc cgagcgtcgg gcaaaggtgt aagactcccg

1901  tcaatagaat gatcagataa ggattaaagg tagttcaata ttaagcttag

1951  cgttagggcc gattagggtt cgaccaaaag cgcaggagac attatgaggt

2001  acttgtagga ggcttgagag gcgcccgcgt atcgatcgag atttgagtga

2051  gattgt
```

FIGURE 4.2.2a) The DNA sequence of *S.equina* mtA-1, the common region and part of the variable region. The mtA-1 coding region is shown underlined. The putative mtA-1 intron, nucleotides 568-624, is shown in plain text. The common region, nucleotides 1579-1636, is shown in italics underlined. The variable region follows on directly after the common region. Primers ordered (Oswel) are shown in italics above the corresponding sequence. Standard pBluescript and PTZ18R primers were used elsewhere. Restriction sites are shown in italics. The putative polyadenlyation signal and CAAT box are shown underlined in upper case.

```
  1  MSGVDQIVKK FANLGEGDRE AAMKAFLAMM PVSNEPVAEP VRKAPTAKKK
 51  VNGFMGFRSN YSPLFSYLPQ KMRSPFMTIL WQYDPYHNEW DFMCSVYSSI
101  RTDLEEQNVT LQLWIHYAIG QMGLIDRDHY MASFGWRLGQ TRNGTTDLFR
151  TAIPMVRRNL QPMNGLCLLI KCLESGLSKH LTNPHPIIAK LQDPSFDMIW
201  INKPPHHQQG HTDQADNSEL GMPSLFPGNH AVAAEVDGIA NLPLSHRTQQ
251  GDFGTEPGFS TQFDTMLDSI LENGNHPSNS HYNMSLAMDL PMTG*
```

FIGURE 4.2.2 b) The deduced amino acid sequence of *S. equina* mtA-1. The region proposed to be a DNA binding domain with similarities to the α domain of *S. cerevisiae* MATα1 and *N. crassa* mtA-1 is shown underlined.

4.2.3 Transformation of *S. equina* mtA-1 into sterile *N. crassa* spheroplasts.

Arnaise *et al* (1993) showed that the *P.anserina* mat+ and mat- mating type genes conferred mating activity when transformed into sterile *N. crassa* mutants, but not vegetative incompatibility or post-fertilization functions. *P. anserina,* unlike *N. crassa* , does not display vegetative incompatibility as opposite mating type nuclei are compartmentalized in a single ascospore. *P. anserina* and *N. crassa* are closely related filamentous ascomycetes. As described in detail in Chapter 1 FMR1 and mtA-1 have each an α domain and FPR1 and mt**a**-1 have an HMG domain.

S. equina and *N. crassa* are closely related and the mtA-1 proteins of both species have an α domain. In equivalent experiments to those of Arnaise *et al* (1993) transformation experiments using the cloned *S. equina* mtA-1 sequence were carried out. Three main questions were to be addressed in these experiments. (i) Could *S. equina* mtA-1 confer mating activity to sterile *N. crassa* mutants? (ii) Could *S. equina* mtA-1 confer homothallism to sterile *N. crassa* mutants that would make them self-fertile? (iii) Would the introduced mtA-1 gene confer the heterokaryon incompatibility function on the recipient mutant?

To answer these questions the mtA-1 gene of *S. equina* was amplified by PCR and cloned into expression vector pBARGRG1. This plasmid contains a glucose-repressible promoter, *grg-1*, and the *bar* gene which gives resistance to Ignite (Pall and Brunelli,

1994). The strategy used for this cloning experiment is outlined in figure 4.2.3a). Figure 4.2.3b) shows a schematic diagram of plasmid pBARGRG1JN.

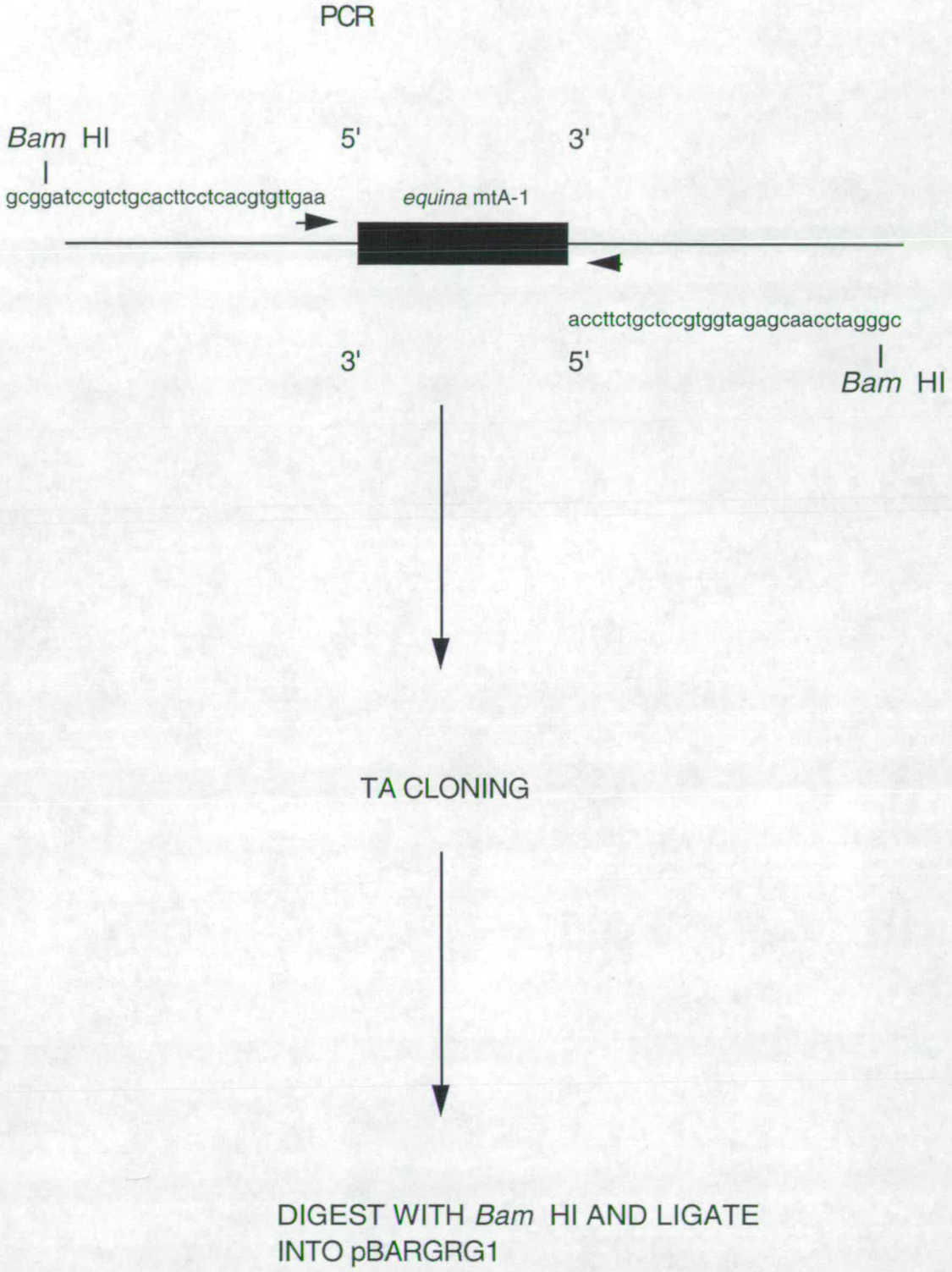
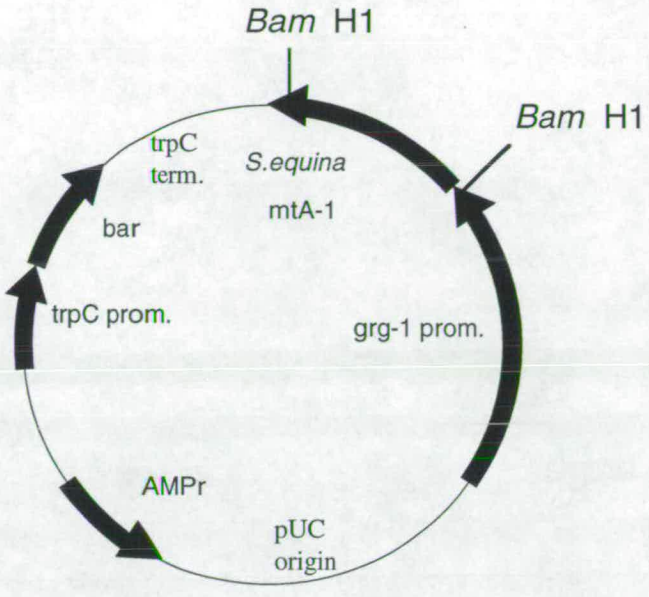


FIGURE 4.2.3a) Strategy for cloning *S. equina* mtA-1 in pBARGRG1. The PCR product, a band approximately 1kb in length was phenol extracted from a low-melting point agarose gel and TA cloned into the pGEM-T Vector. The plasmid was digested with *Bam* HI to cut out the mtA-1 gene, the fragment extracted from the gel and ligated into pBARGRG1 and the plasmid transformed into *E. coli* cells.



pBARGRG1JN
7.35 kb

FIGURE 4.2.3b) Schematic diagram of plasmid pBARGRG1JN. The orientation of the *S. equina* mtA-1 gene was checked using a unique *Eco* RI sight from the polylinker and a unique *Sal* I sight from the mtA-1 gene inton, near the N-terminus of the gene. For the MtA-1 gene to be in the correct orientation then a *Eco* RI/*Sal* I digest should give a band 0.7 kb in size.

To test the ability of *S. equina* mtA-1 to confer mating activity onto sterile *N. crassa* mutants, the pBARGRG1JN plasmid was transformed by electroporation into *N. crassa* a^{m1} spheroplasts. a^{m1} mutants have a 1-base deletion causing a frameshift in the mt**a**-1 coding region resulting in a loss of mating and vegetative incompatibility functions. Electroporation of the spheroplasts is described in 2.2.3.2. As a control in this experiment some spheroplasts were electroporated without pBARGRG1JN DNA. The transformed spheroplasts were allowed to regenerate their cell walls and then plated onto selective ignite medium. Two transformants grew on the plate onto which conidia electroporated with pBARGRG1JN had been plated. The control plate also had a colony growing on the selective medium, this colony does not contain pBARGRG1JN and therefore could not be resistant to Ignite. Genomic DNA was extracted from the three transformants, digested with *Eco* RI and ran on an agarose gel along with *Eco* RI digested genomic DNA from *N. crassa* mtA and λ *Hind* III size markers. The gel was Southern blotted and probed with *N. crassa* mtA probe to look for the integration of the pBARGRG1JN plasmid. Figure 4.2.3c) shows the result of this experiment.

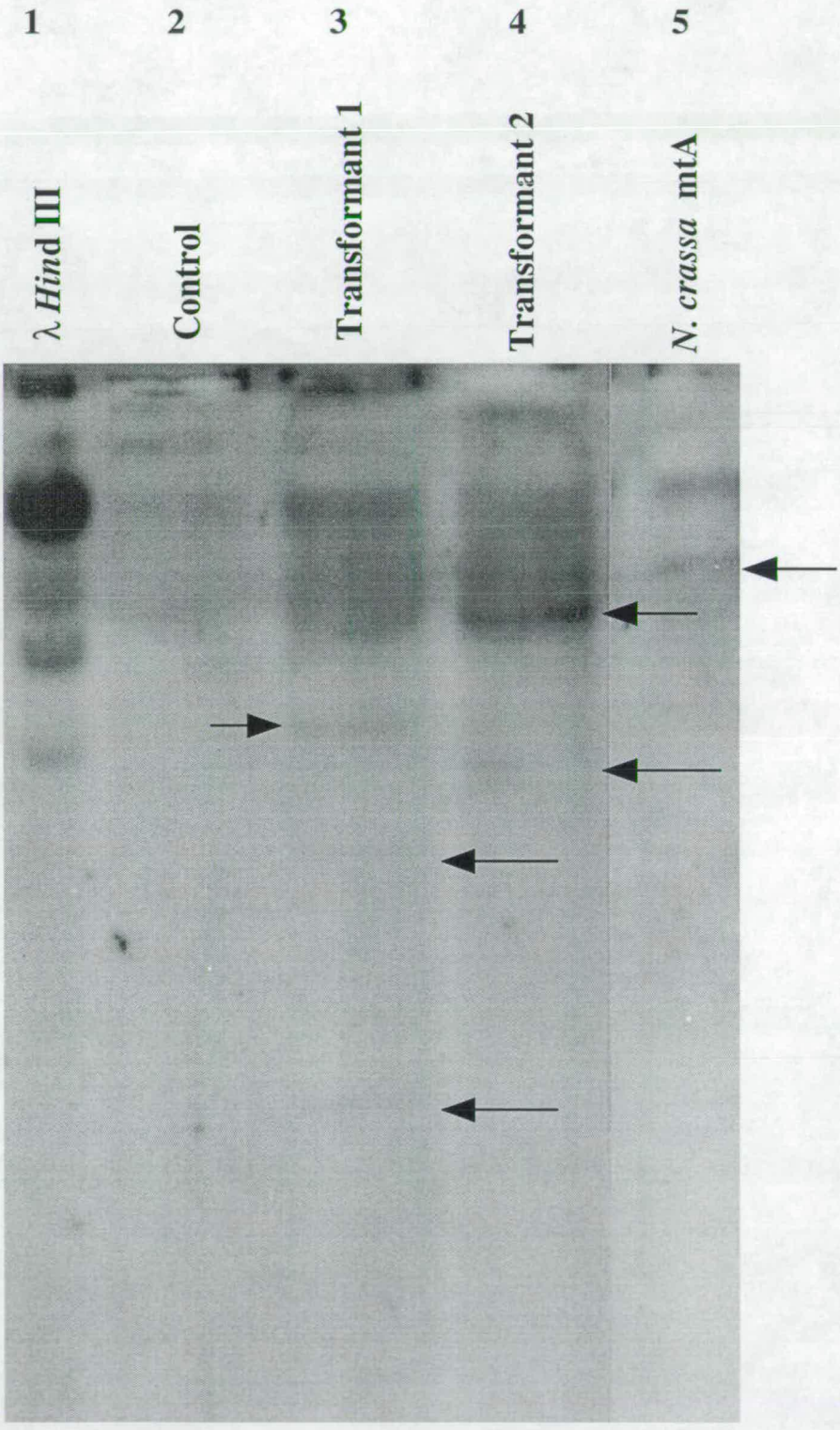


FIGURE 4.2.3 c) Southern blot of genomic DNA extracted and digested with *Eco* RI from the Control, Transformant 1, Transformant 2 and *N. crassa* mtA. The blot was probed with the *N. crassa* mtA probe and a λ probe. Fragments that hybridised with the mtA probe are indicated.

The genomic Southern blot shown in figure 4.2.3c) demonstrates that the *N. crassa* mtA probe did not hybridise to the digested genomic DNA from the control sample. Bands corresponding to the pBARGRG1JN plasmid are only observed in transformant 1 and transformant 2. The bands in lane 5 correspond to mtA and not the pBARGRG1 plasmid itself. Clearly the control colony growing on Ignite containing medium was a spontaneous resistant.

Transformants 1 and 2 and the control transformant were crossed to *N. crassa* mt**a** to observe whether plasmid pBARGRG1JN conferred mating activity to the transformants. Wild type *N. crassa* mtA was crossed to mt**a** as a control. Figure 4.2.3d) shows the result of these crosses.

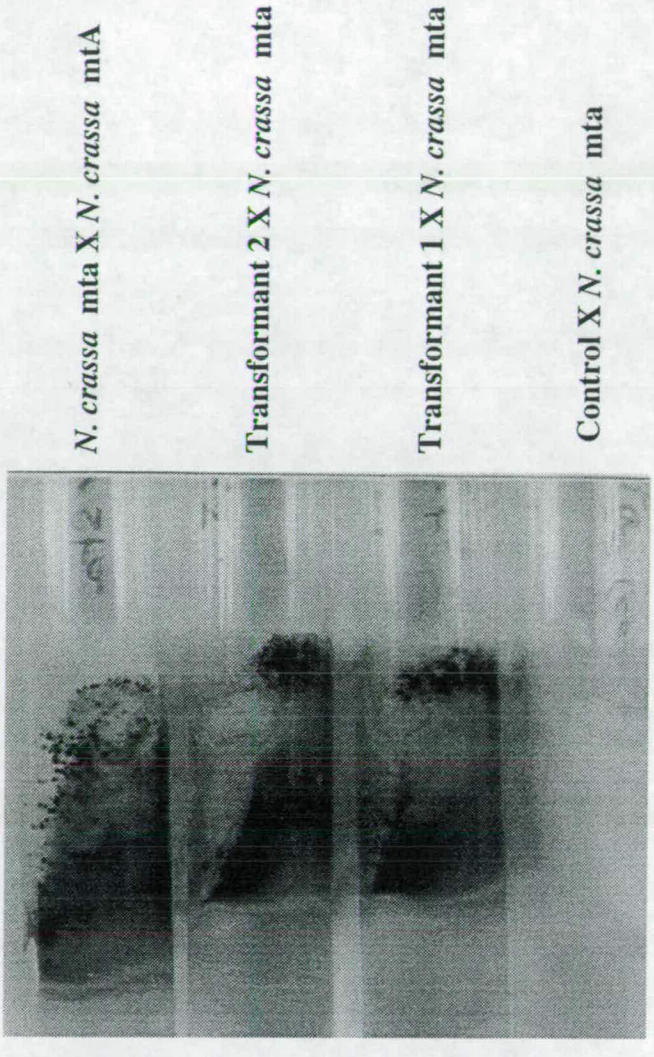


FIGURE 4.2.3 d) The result of crossing transformants 1 and 2 and the control with *N. crassa* mt**a.** A wild type cross was included in the experiment and one can see perithecia were formed as a result of this cross. The crosses involving transformants 1 and 2 resulted in perithecia being formed. No perithecia were produced in the control cross.

From figure 4.2.3d) one can see that in the wild type cross perithecia are formed indicating that mating has taken place. The crosses involving transformants 1 and 2 also show the presence of perithecia. The plasmid pBARGRG1JN containing the *S. equina* mtA-1 gene has conferred mating type activity on the sterile *N. crassa* mta mutant. The cross involving the control resulted in no perithecia being formed.

A second transformation experiment was undertaken as a further control to show that it is expression of the *S.equina* mtA-1 gene that is conferring mating type activity to the sterile *N. crassa* mt**a** mutant, not the expression vector itself. The plasmid pBARGRG1JN was transformed once again into the a^{m1} spheroplasts as was plasmid pBARGRG1, the expression vector without the *S. equina* mtA-1 gene. Transformants were plated onto the selective ignite medium and pBARGRG1 and pBARGRG1JN transformants crossed to *N. crassa* mt**a** and mtA. Figure 4.2.4e) shows the result of these crosses.

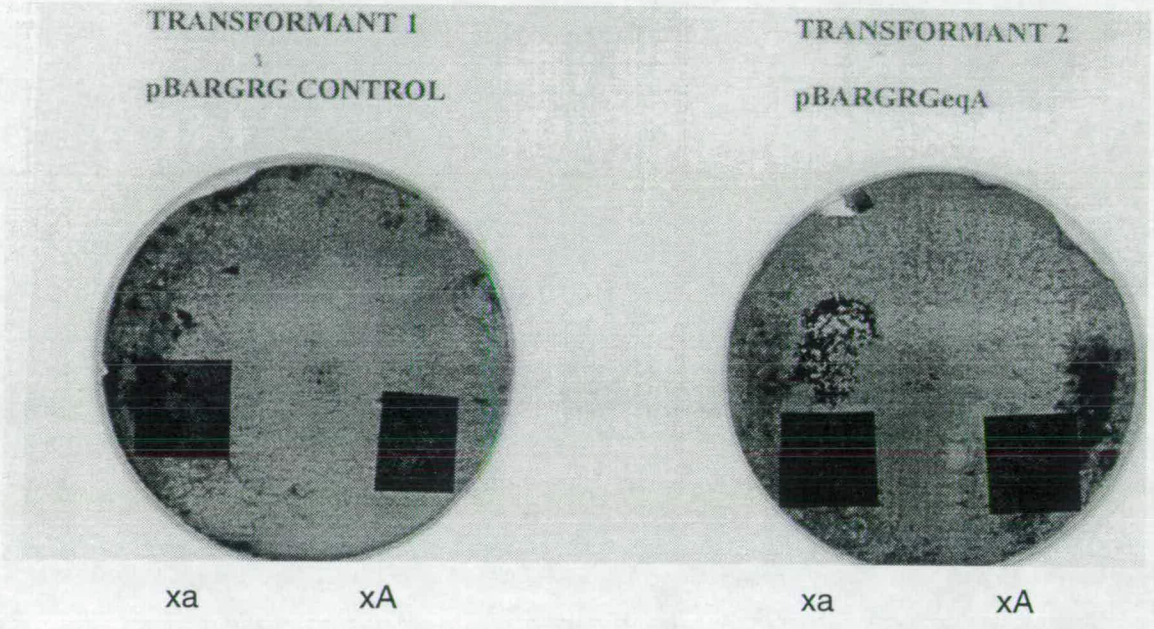


FIGURE 4.2.3 e) The results of crossing transformant pBARGRG, a transformant containing the plasmid without the *S. equina* mtA-1 gene, and transformant pBARGRGeqA with *N. crassa* mtA and mt**a**. No perithecia are formed on the plate where pBARGRG was crossed to both mating types. On the plate where pBARGRGeqA was used in the crosses, no perithecia are produced in the mtA cross but perithecia are produced in the mt**a** cross. This indicates that the *S. equina* mtA-1 gene is conferring the ability to mate on the sterile transformant.

Figure 4.2.3e) shows the formation of perithecia only on the plate where pBARGRG1JN (labelled pBARGRGeqA in figure 4.2.3e)) was crossed to *N. crassa* mt**a**. No perithecia were produced when pBARGRGeqA was crossed to mtA. For the control crosses involving pBARGRG1, pBARGRG1 x mtA/mt**a**, no perithecia were produced as a result of either of these crosses. *S. equina* mtA-1 confers mating type activity when transformed into a sterile *N. crassa* mt**a** mutant.

pBARGRG1JN transformants were plated onto crossing medium (Westergaard and Mitchell, 1947) and examined for homothallic behaviour. No perithecia were formed under these conditions.

Examination of perithecia produced by the pBARGRG1JN x *N. crassa* mt**a** cross revealed that no asci or ascospores were observed as a result of this cross.

The original *N. crassa* mating type mutant used in these transformation experiments had, in addition to a mutation at the mating type locus, a requirement for adenine brought about by a mutation at the *ad-3B* locus. It should have been possible to test the *Neurospora* transformants to observe whether or not the *S. equina* mtA-1 sequence had conferred vegetative incompatibility function as well as mating type function. An experiment was set up in which the *ad-3B* mutation was used as a forcing marker to see if heterokaryons could be formed with *pan-1 al-2* strains of either mating type. The latter strains were chosen so that they were compatible at all other heterokaryon incompatibility loci so any incompatibility between the transformants and tester strains would be due to the *S. equina* mtA-1 sequence. Both the transformants and the non-transformed controls unfortunately were found to be prototrophic and showed no sign of adenine requirement. This meant the incompatibility function of the *S.equina* mtA-1 could not be assessed in the *Neurospora* transformants. A small scale experiment to try and recover the *ad-3B* mutation from the strains, using the fact that the mutation confers a purple phenotype on mutant colonies, was not successful and therefore the experiment was abandoned.

4.3 Discussion

The *Sordaria equina* mtA-1 gene, using the definition of mtA-1 by Glass *et al* (1990a), comprises of two exons separated by an intron 57bp long which is found 176bp downstream of the first ORF ATG start site. The position of the intron is conserved in relation to the *N. crassa* mtA-1 intron (see chapter 6). The intron has a 5' splice site GTAAGT and 3' splice site CAG. A putative CAAT box is located at nucleotide 220 and a putative polyadenylation signal located at nucleotide 1474. If the putative intron were spliced out, the first and second exons would code for a translational product encoding a 294 amino acid polypeptide.

The presence of a second intron in *S. equina* mtA-1 is more difficult to determine from the sequencing data. The 5' splice sequence is conserved (GTTAGT, nucleotides 1330-1335 figure 4.2.2a)) but no 3' splice sequence appears to be conserved after 60bp. A 3' splice sequence CAG is found 75bp downstream of the 5' splice sequence but no in-frame stop codon exists after this CAG. When the mtA-1 gene was amplified by PCR for expression studies the 3' primer used was taken from the proposed second intron and therefore did not include any of the additional amino acids which would have been present if the second intron was spliced out.

The expression studies of *S. equina* mtA-1 have demonstrated that this gene encodes a functional polypeptide. This is interesting considering that *S. equina* does not mate with a strain of opposite mating type as only one mating type sequence (A) is present in the genome of the species. Introducing *S. equina* mtA-1 into *N. crassa* did not confer homothallism. Likewise when *N. africana* mtA-1 was introduced into *N. crassa,* homothallism was not conferred (Glass and Smith, 1994). It is likely that differences elsewhere in the genome of a homothallic species are responsible for homothallic behaviour. Glass and Smith (1994) performed mRNA analysis to obtain evidence for mtA-1 transcription in *N. africana* itself. This experiment was not performed for *S. equina* but as the cDNA of mtA-1 was detected in *N. africana*, it seems likely that mtA-1 is expressed in *S. equina.* Expression of mtA-1 in homothallic species might reflect the need for the gene for post-fertilization functions.

Transformation of the mtA-1 gene for the expression study of *S. equina* mtA-1 resulted in ectopic integration of the gene into the *N. crassa* genome because the bands in the Southern blot shown in figure 4.2.3c) were not the same as for wild type *N. crassa* mtA. The *S. equina* mtA-1 gene confers the ability to mate upon the recipient but no

ascospores are produced as a result of this mating. The obvious answer to this observation is that one is only transforming mtA-1 into *N. crassa* sterile mt**a**. MtA-2 and mtA-3 are required for post-fertilization functions also in *N. crassa* mtA (Ferreira *et al*, 1996). However ectopic integration of *N. crassa* mt**a**-1, a gene which has mating and post-fertilization functions, into mtA strains suppressed for vegetative incompatibility results in mt**a** mating type activity being conferred but no ascospore production when the transformants are crossed to mtA (Staben and Yanofsky, 1990). Chang and Staben (1994) showed that if mtA was directly replaced by mt**a** DNA in *N. crassa*, effecting a mating type switch, then this mt**a** strain when crossed to mtA will produce ascospores. Therefore mating type DNA may function properly only when it is present at the correct chromosomal location.

Chapter 5

The Sub-Cloning and Sequencing of *Sordaria fimicola* mtA-1 and mta-1

5.1 Introduction

Sordaria fimicola is a homothallic species with morphological similarities to *Sordaria equina*. With heterothallic strains it would be customary to try and cross two strains to deduce if they belong to one species or are two distinct species. With homothallic species being self-fertile this is not possible. Figure 5.1 shows a genomic Southern blot where genomic DNA from *N. crassa* mt**a** and mtA, *S. fimicola* and *S. equina* has been digested and probed with *N. crassa* mt**a**. This result is consistent with the taxonomic conclusion that *S. fimicola* and *S. equina* are two distinct species.

Bands indicating hybridisation are clearly seen in lanes 1, 2 and 3 where digested *N. crassa* mt**a** genomic DNA is present. Bands are also present in lanes 4 and 5 indicating that *S. fimicola* hybridises with the mt**a** probe. No hybridisation with the mt**a** probe is seen for *S. equina*, lanes 7, 8 and 9. Some background signal is present in the *N. crassa* mtA digests, lanes 10, 11 and 12 but the strong signals obtained from the *N. crassa* mt**a** digests are satisfactory as a control. No bands are seen in lane 6, *S. fimicola* genomic DNA digested with *Xho* I. The genomic DNA seemed resistant to digestion with *Xho* I in this case hence no bands are distinct. A 4.7kb *Bam* HI band is highlighted, this fragment was present in a λ EMBL3 clone obtained from the honours cloning practical. Sub-cloning of this fragment and others is described in 5.2. The plaque from which the λ clone was obtained hybridised to both mtA and mt**a**.

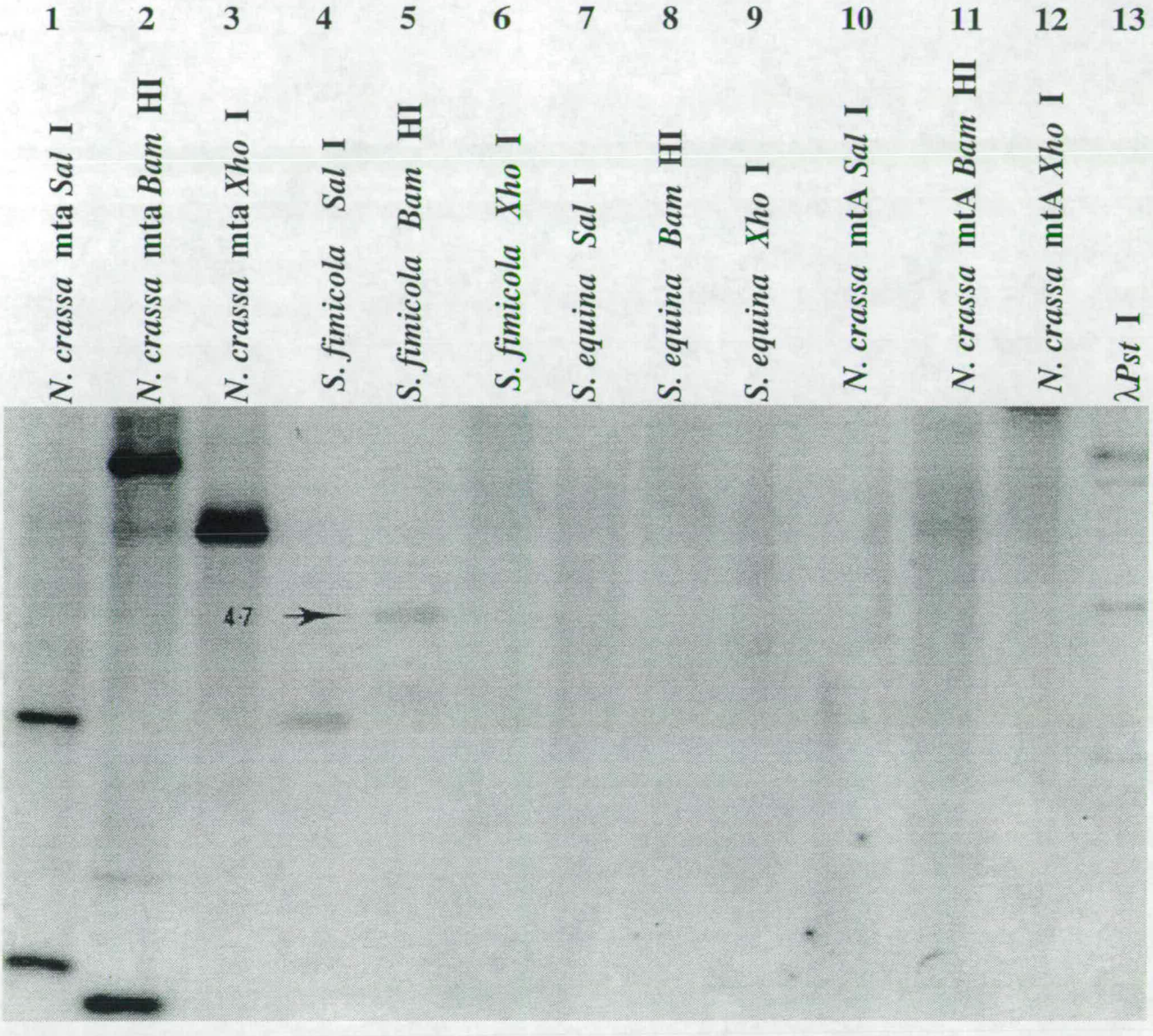


FIGURE 5.1

Genomic Southern blot of DNA exctracted and digested from *N. crassa* mt**a** and mtA, *S. fimicola* and *S. equina.* The blot was probed with the *N. crassa* mt**a** probe. Enzymes used are indicated above the figure. The 4.7 kb *Bam* HI fragment which hybridised with the probe is indicated also.

5.2 Results

5.2.1 Sub-cloning of *S. fimicola* mtA-1 and mta-1.

Figures 5.2.1a) and b) show a Southern blot of digested *S. fimicola* λ EMBL3 DNA probed with mtA (a) and mt**a** (b). Figure 5.2.1c) shows the restriction map of the *S. fimicola* λ EMBL3 clone isolated using the *N. crassa* mtA and mt**a** probes. The sub-cloning strategy is shown also.

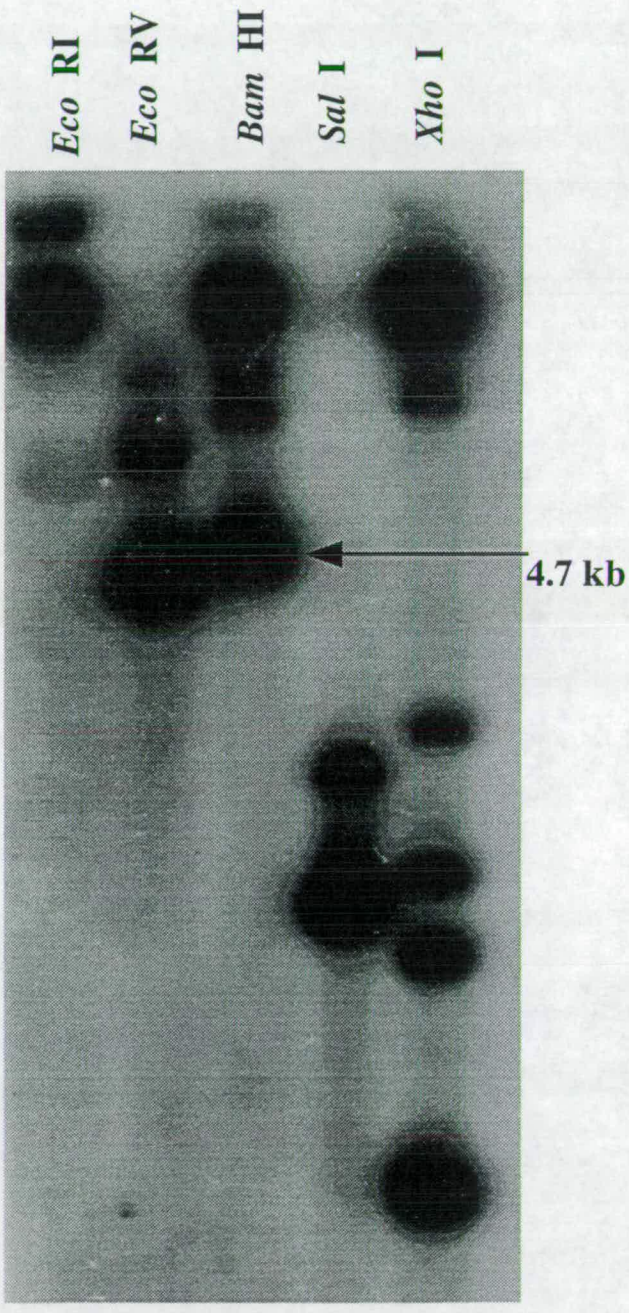


FIGURE 5.2.1 a)

Southern blot of *S. fimicola* λ DNA digest probed with mtA probe. Enzymes used are indicated above the figure. The 4.7 kb band which hybridised with the mtA probe is indicated also.

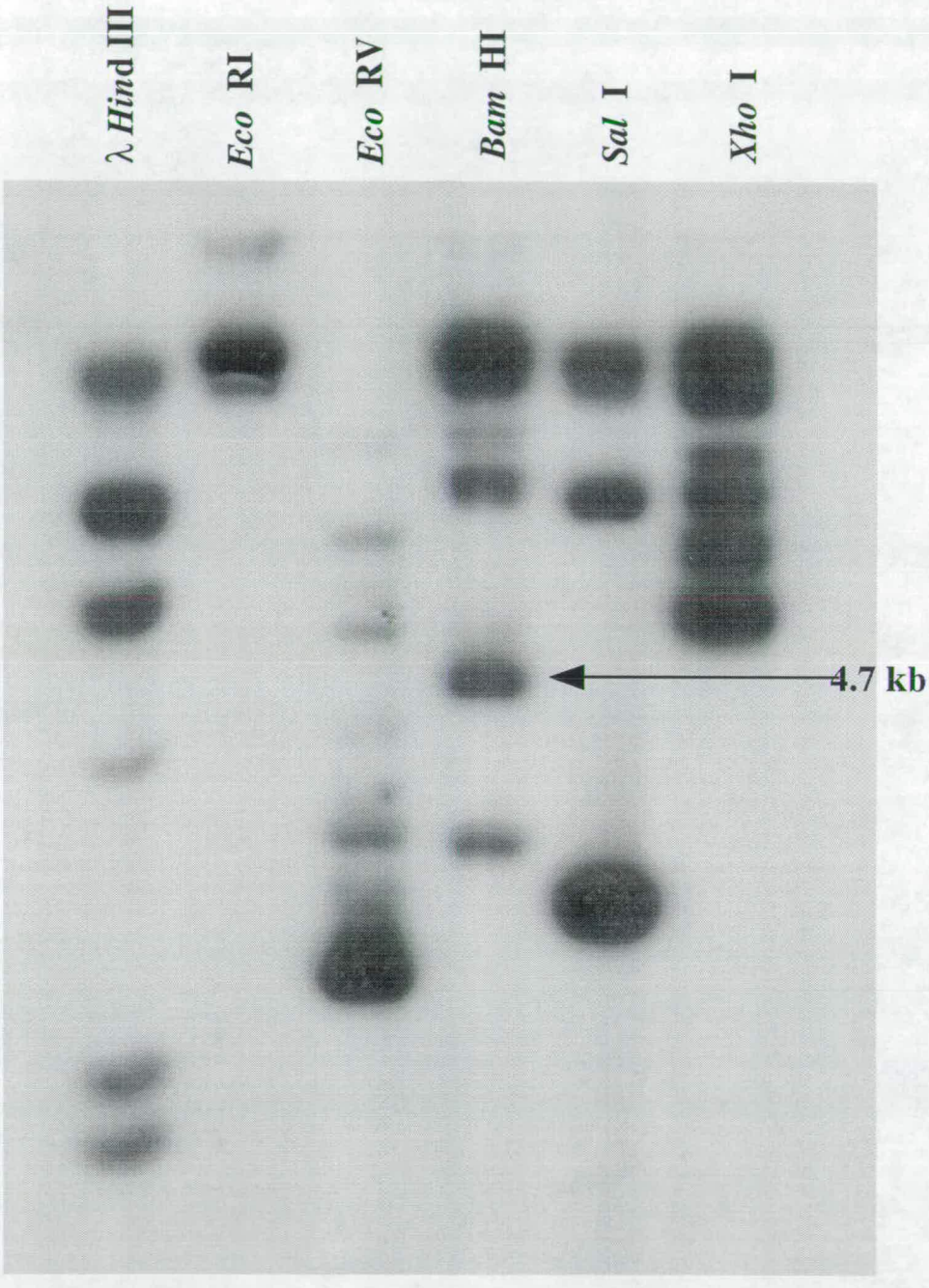


FIGURE 5.2.1 b)

Southern blot of *S. fimicola* λ DNA digest probed with *N. crassa* mt**a** probe and λ probe. Enzymes used are indicated above the figure. The 4.7kb band which hybridised with the mt**a** probe is indicated also.

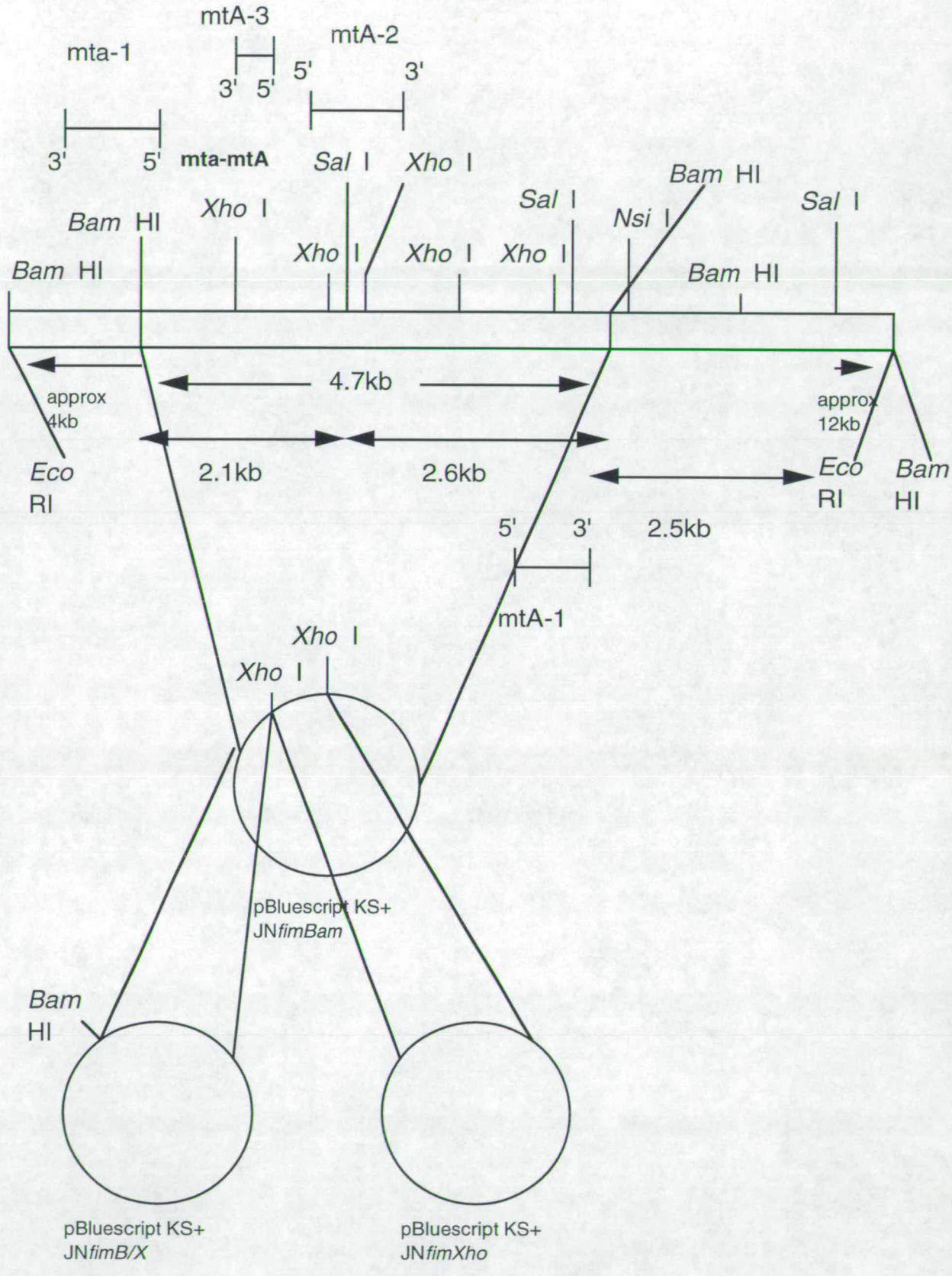


FIGURE 5.2.1c) Restriction map and sub-cloning strategy for *S. fimicola* λ EMBL3. The positions of the mt**a**-1 and mtA-1, mtA-2 and mtA-3 ORFs are shown as is the location of the mta-mtA junction.

The two blots photographed in figures 5.2.1 a) and b) were superimposable and show that the 4.7kb *Bam* HI fragment hybridises with both mtA and mt**a**. A partial digest can be seen for the *Eco* RV and *Xho* I digests in figure 5.2.1b). Blot a) was stripped and reprobed with mt**a** and a λ probe. *S. fimicola* mtA-1 and mt**a**-1 are linked.

Figure 5.2.1 c) describes the sub-cloning strategy so that mtA-1 and mt**a**-1 of *S. fimicola* could be sequenced. Most of the of mt**a**-1 gene was not present on the 4.7kb *Bam* HI fragment. A 4kb *Bam* HI fragment whch hybridised to the mt**a** probe only (see figure 5.2.1 b)) was subcloned from the λ clone, this fragment is adjacent to the 4.7kb *Bam* HI fragment on the clone and contains the rest of the mt**a**-1 coding region. Plasmids JN*fimB/X* and JN*fimXho* were constructed to sequence the junction between the mtA and mt**a** *S. fimicola* idiomorphs. The 2.1kb *Sal* I/*Bam* HI fragment was sub-cloned also and from this sequence data from mtA-2 was obtained. An *Nsi* I site is present on the clone, this is most likely to indicate the presence of the mtA common region (Randall and Metzenberg, 1995) although this region of the clone would need to be sequenced to confirm this.

5.2.2 Sequencing of *S. fimicola* mtA-1 and mta-1.

Double-strand sequencing was performed on DNA prepared by the boiling method. Automated sequencing was performed on DNA prepared by alkaline lysis. Primers used were either standard pBluescript KS+ primers or primers ordered from Oswel which annealed to the mtA or mt**a** sequences. Figure 5.2.2a) shows the DNA sequence of mtA-1. Figure 5.2.2b) shows the deduced amino acid sequence of mtA-1. Figure 5.2.2c) shows the DNA sequence of mt**a**-1 and figure 5.2.2d) shows its deduced amino acid sequence. Figure 5.2.2e) shows upstream sequences and part of the coding sequence of MtA-2.

```
   1  atgtccagcg tcgatcaaat cgtcaagacg ttcgccaacc tccctgaggg

  51  cgagcgcaac gcagcagtca atgctatctt agccatgatg ccccccggcc

 101  ctggtcctgt tcgccaaatc cccgaacctg ttccacaagc ccccgcgcca

 151  aagaagaagg tcaacggctt catgggtttc agatgtaagt caaatctgaa
                                                        3' gat
 201  ttaatcttga cgacgatcca tactgattgc ctttctattt cagcgtacta
      aaggagcgag aagag  5'
 251  ttcctcgctc ttctctcagt ttcctcagaa ggcgcgatcg cccttcatga

 301  ccatcctctg gcagcacgat ccctttcaca acgaatggga tttcatgtgc
                                      Xho I
 351  tcggtgtatt cgtcaatccg caactacctc gagcagttga acgcgcagcg

 401  ggagaagaag attaccctgc aatactggct tcactttgct gtccccgtca

 451  tgggagtgct tggtcgcgaa aactacttgc ccacgcttgg ctgggacctc

 501  gtcacgatgc ccaacggcac tatcgacctt atgcgcatcg ctatgccttt

 551  gtttagaaag aacctccagc ccatggacgg cctatgcctg ttcaccaagt
                            Sal I
 601  gtcaggaggg cggattgcaa gtcgacaacc agcacttcgt cattgccaag

 651  ctttcagatc ctagccacga catgatctgg ttcaacaagc gccctcacta
          3'ctct gtgcggcggg tttg  5'
 701  tcagcagaga cacgccgccc aaaccgacag ttctgaactc ggtgtgtcgg

 751  cgctcttccc tcgcaatcac gcagttgctg cagaggcaga tggcgtcgcc

 801  actgttcaac tccctcattg gatgcagcag ggagatttcg gcaccgagtc

 851  cggatactca cctcagtttg agaccttgtt gggttccata cttgagaatg

 901  gaaacgccac cagtaatgat tcctacaaca tggctctggc tatggatgtt

 951  cctatgatgg gttagtggat gatgaagtgc catgtcactt agctttacta

1001  gtgtgctgac gatttggcag gattcaatgg aggagcatag aagtacggca

1051  cagtcacaac tttcctttcc tttcctttgt caaatctggt ttcgtggtat

1101  gtgcatacaa agcgatggcg aaaagggtct agttaggttt ctttgtgcat

1151  tcattcga
```

FIGURE 5.2.2a) The DNA sequence of *S.fimicola* mtA-1. The coding region is shown underlined. The putative mtA-1 intron, nucleotides 185-243, is shown in plain text. Restriction sites are shown in italics. Primers ordered (Oswel) are shown in italics above the corresponding sequence.

```
  1  MSSVDQIVKT FANLPEGERN AAVNAILAMM PPGPGPVRQI PEPVPQAPAP
 51  KKKVNGFMGF RSYYSSLFSQ FPQKARSPFM TILWQHDPFH NEWDFMCSVY
101  SSIRNYLEQL NAQREKKITL QYWLHFAVPV MGVLGRENYL PTLGWDLVTM
151  PNGTIDLMRI AMPLFRKNLQ PMDGLCLFTK CQEGGLQVDN QHFVIAKLSD
201  PSHDMIWFNK RPHYQQRHAA QTDSSELGVS ALFPRNHAVA AEADGVATVQ
251  LPHWMQQGDF GTESGYSPQF ETLLGSILEN GNATSNDSYN MALAMDVPMM
301  G*
```

FIGURE 5.2.2b) Deduced amino acid sequence of mtA-1. The region proposed to be a DNA binding domain with similarities to the *S. cerevisiae* α domain is shown underlined.

```
  1  atggaaaaca acttgatgca ccccgctcgg acgtcagcgg aactcagggt
 51  caccatggct tggtctggca tctcgaacca gcttgggcac tggaacgacc
101  gcaagatcat tgccattcct ctgagtgact tcactatcgc ccaccctgac
151  attcatgctg gcatcgtcgc cgaatacaag taagtgtcct cacccatctc
201  tcaccttacc ttatactgac catttgcatt aggaaagcga ctggtgaaga
251  gggcatgttt gctcgcgata ccgaggcact ggaaatcatg ctgcttggcc
301  ccgccaagct gtttaaagcc gatagtgtcg ttgttgagag caatctgttt
       Bam HI
351  tgggatccca agggcatcca tgctgagaca cctaagcagc agcagaagaa
401  gaaggccaag attcctcgtc cgcccaatgc ctacatcttg taccgcaagg
451  accatcatcg tcagatccgc gagcagaacc ctggactgca caacaatgag
501  atctgtaggt ctcttgtcac tatgatctat attggttgac cctgagacta
                                  caacat gtggcgtgat ga
551  acctcactta gcggtcattg ttggcaacat gtggcgtgat gagcagccgc
601  acattcgcga caaatatttc agcatggcca atgaggtcaa ggctagattg
651  ttgctggaca atcccgacta tcgctacaat gcccgtcggt ctcaggacat
701  tcgcaggcgc gtttcgccgt atctcaagat caagctcctc aattatgacg
751  tcaacggcaa ccttctttgg ggcaccgtca acgccgagga tgccgcgcta
```

```
 801  atccgcactc actttcatgg agtcgtccgt gtcgaggaaa cggacgaagg

 851  ctgcagaatt gtctgccgtc ccgtcgcagg atctaggaag cttcgcgccg
                                                        cccgtc
 901  ccaacgtcga cacatggatg cctcgctata cggttgatgc taatcccgtc
      tctgaggaag ac
 951  tctgaggaag acgaggcagc ccagggtacg ctctttaact tcaacgacac

1001  cttggacggt ttctttccca tgaatgaaca cctctggggc accgctaccc

1051  aaaacccttc ccccgctctc gccaacatcc acgcactcct ggattttggt

1101  caccccaaca gcgtgcaggc aatcacccaa aacattcaga acatgattac

1151  ccaggtccat atcatgacgc tgccgctgct gccacctctg cgcctgccgt

1201  caatctggtc atggctgacg ataccatcaa cccagcgctc attcccactg

1251  tga
```

FIGURE 5.2.2c) DNA sequence of *S. fimicola* mt**a**-1. The coding region is shown underlined. The putative introns, nucleotides 180-232, 505-561, are shown in plain text. Restriction sites are shown in italics. Primers ordered are shown in italics above the corresponding sequence.

```
   1  MENNLMHPAR TSAELRVTMA WSGISNQLGH WNDRKIIAIP LSDFTIAHPD

  51  IHAGIVAEYK KATGEEGMFA RDTEALEIML LGPAKLFKAD SVVVESNLFW

 101  DPKGIHAETP KQQQKKKAKI PRPPNAYILY RKDHHRQIRE QNPGLHNNEI

 151  AVIVGNMWRD EQPHIRDKYF SMANEVKARL LLDNPDYRYN ARRSQDIRRR

 201  VSPYLKIKLL NYDVNGNLLW GTVNAEDAAL IRTHFHGVVR VEETDEGCRI

 251  VCRPVAGSRK LRAANVDTWM PRYTVDANPV SEEDEAAQGT LFNFNDTLDG

 301  FFPMNEHLWG TATQNPSPAL ANIHALLDFG HPNSVQAITQ NIQNMITQVH

 351  IMTLPLLPPL RLPSIWSWLT IPSTQRSFPL *
```

FIGURE 5.2.2d) Deduced amino acid sequence of *S. fimicola* mt**a**-1. The region thought to be a DNA binding domain with similarity to the high mobility group (HMG) proteins is shown underlined.

```
  1  gtcaaggact tctgccaacc tcacgggaga taacaagaac ttggacaaaa
 51  tttggcatgc aataccccac aggtcaagtg atacccgccc caagtctcac
101  aggagaacaa taggacggac ctgagattga aaccgacacg caagccgccc
151  gaaagagttt gggaacctga gtccttaagt tgttcacttc ttcgaaatgt
201  cctgtcggca aaacttctcc tgctacccag cggtcttctt gtagcttcct
                                                    Xho I
251  cgagcactca ctgttgagtc ttccttttcg ttagcgaccg acatggagct
301  catcaacacg caacgcacct cgggccagaa aggacaggat ctcgaaatgg
351  tgtacaaggt aacaacatgt ctaccctcga tacactcatt tacttatcgc
401  tgatgaactg gccagaaact ccatcagtta caggctaggc tttctcgttc
451  atatctttca gaggcaatca aggagttcga agagaacctt cagtgtcttt
501  ttcatgaagc caagatcttg ctatgcacga aaagaacgaa gtatcgccaa
551  agctggtttg ggtctagcaa cgaggtcggg cctaacaacg aggaaaaaat
601  catcaaggca gcatgctgcg
```

FIGURE 5.2.2e) DNA sequence of upstream regions and part of the coding region of *S. fimicola* mtA-2. Coding sequences are underlined. The first intron in the gene (one of four), nucleotides 359-415, is shown in plain text. Restriction sites are shown in italics.

5.2.3 Discussion

The *Sordaria fimicola* mtA-1 gene, using the definition of mtA-1 by Glass *et al*, (1990a), comprises of two exons separated by an intron 59 bp long which is located 185 bp downstream of the first ORF ATG start site. The intron has a 5' splice site GTAAGT and 3' splice site CAG. The position of the intron is conserved in relation to the position of the intron in *N. crassa* mtA-1 (see chapter 6). If the putative intron was spliced out the first and second exons would code for a translational product 301 amino acids long.

The 5' splice sequence of the proposed second intron (Saupe *et al* , 1996) is conserved in *S. fimicola* mtA-1 (GTTAGT, nucleotides 961-966 figure 5.2.2 a)) as is the 3' splice sequence (CAG, nucleotides 1018-120 figure 5.2.2 a)). A putative second intron 60bp long exists in the mtA-1 gene. If this intron were spliced out then five amino acids would

be added to the *S. fimicola* mtA-1 ORF making it a 306 amino acid long polypeptide with a stop codon, TAG, at nucleotide 1040 in figure 5.2.2 a).

The *Sordaria fimicola* mt**a**-1 gene comprises of three exons separated by two introns. The first intron is 53bp long situated 180bp downstream of the first ORF ATG site. The second intron is 57bp long and 505bp downstream of the first ORF ATG site. The first intron has a 5' splice site GTAAGT and 3' splice site TAG. The second intron has a 5' splice site GTAGGT and a 3' splice site TAG. The positions of both introns are conserved in relation to the position of the introns in *N. crassa* mt**a**-1 (see chapter 6). If the putative introns were spliced out the first, second and third exons would code for a translational product 380 amino acids long.

The *Sordaria fimicola* mtA-2 gene was only partially sequenced but the first intron is shown, 57bp long with a 5' splice site GTAACA and 3' splice site CAG. The intron is situated 67 bp downstream of the first ORF ATG site. The position of the intron is conserved in relation to the position of the first intron of *N. crassa* mtA-2. Upstream sequences of *S. fimicola* mtA-2 are shown, these are proposed to be important in transcriptional and translational control of *N. crassa* mtA-2 and mtA-3 (Ferreira *et al*, 1996). The upstream sequences of *S. fimicola* and *N. crassa* are compared in chapter 6.

The *Sordaria fimicola* mtA-1 and mt**a**-1 genes are linked. JN*fimB/X* and JN*fimXho* were constructed so the junction between the two genes could be sequenced (see figure 5.2.1c)). The *Xho* I end of plasmid JN*fimB/X* was sequenced and corresponded to region approximately 0.9kb upstream of *N. crassa* mt**a**-1. One of the *Xho* I ends of plasmid JN*fimXho* should follow on directly from the sequencing data obtained from JN*fimB/X* as both fragments lie adjacent to each other. Both ends of JN*fimXho* were sequenced, one end corresponded to the *Xho* I site in MtA-2 seen in figure 5.2.2e) at nucleotide 296. The other end followed on from JN *fimB/X* for 17 nucleotides in mt**a** and then proceeded into a region of *N. crassa* mtA-3, approximately 0.4kb downstream of the ATG start site of the gene.

The mtA-2 and mtA-3 genes are transcribed divergently to each other as shown in figure 5.2.1c). When the mtA-3 end of JN*fimXho* was sequenced, the direction of sequencing was 3'-5' with respect to mtA-3 when the data was compared with the *N. crassa* mtA-3 sequences. The junction between *S. fimicola* mt**a**-1 and mtA-3 cuts out approximately 0.8kb of the downstream portion of the mtA-3 ORF. Beatty *et al* (1994) probed genomic DNA from *N. terricola* , a homothallic species containg mtA and mt**a**, with probes A2-

A6 and **a**1-**a**5 as described in section 1.2.3.2. The A2 probe, 0.8kb long, covers part of mtA-3. This probe did not hybridise with *N.terricola* genomic DNA indicating that this species like *S. fimicola* is missing part of the mtA-3 gene. Probe **a**4, 0.4kb long, covers a region adjacent to the right flank upstream of the mt**a**-1 gene. **a**4 does not hybridise with *N. terricola*. The position of the mt**a**/mtA junction in *S. fimicola* means that this species is also missing a substantial portion of the the region upstream of mt**a**-1, approximately about 1kb. *S. fimicola* appears to be missing the mt**a** common region and variable region due to the junction between mtA and mt**a**.

Figure 5.2.3a) shows the *S. fimicola* mtA-mt**a** junction in relation to its location at *N. crassa* mtA and mt**a.**

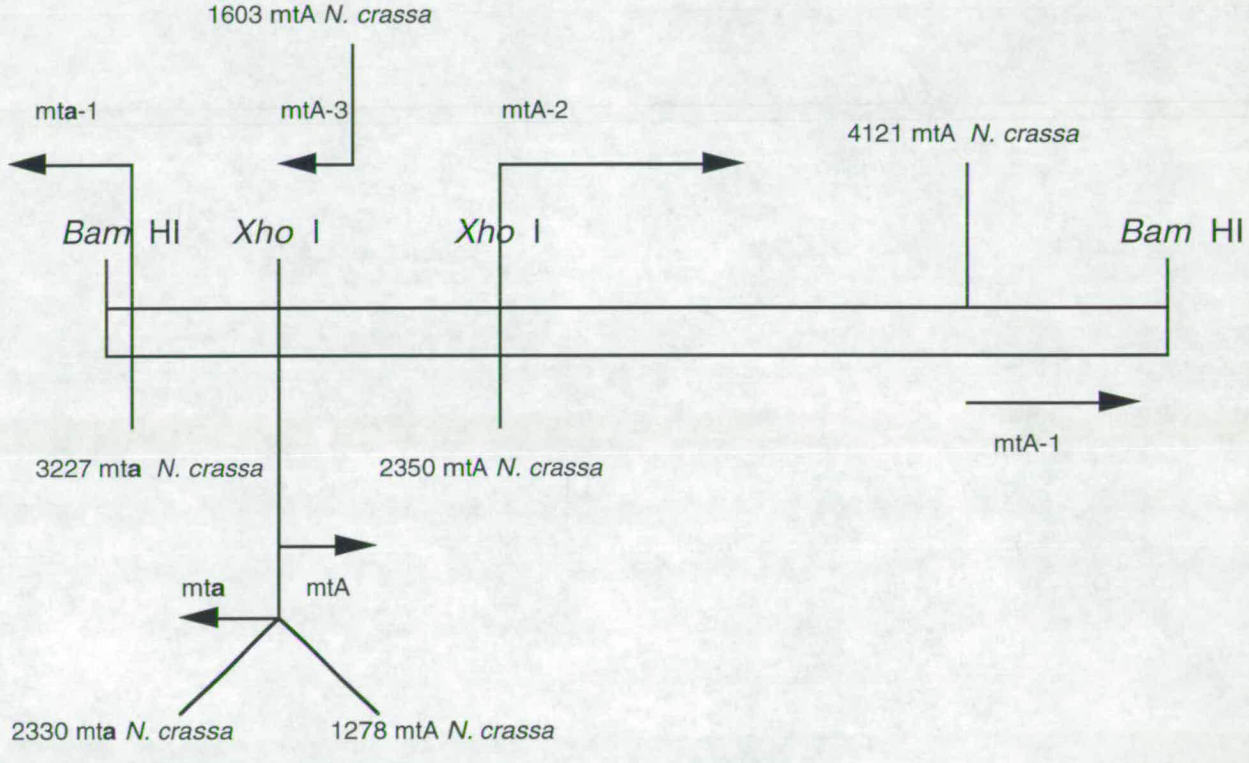


FIGURE 5.2.3a) Schematic diagram of the mtA-mt**a** junction of *S.fimicola* in relation to mtA-1, mtA-2, mtA-3 and mt**a**-1. Numbers shown represent the equivalent position in *N. crassa*.

Figure 5.2.3b) shows the sequence of the junction of mtA/mt**a** in *S. fimicola* compared to the respective sequences in *N. crassa.*

mta mtA-3

TGTTCGGCTCTCACAGAGCTCTTCCGTAAGG/GACACTCGTACAGTGTCAAAGGGGTAAGG
S. fimicola

TATTTGGTTCTCACAGAGCTCTTCCATCAGG*ATCGAGCATCTGCCTTGGAGCTTCTCTCGTCTCGCTTT*
N. crassa mt**a**

*CCCCCGGGTTACAAACCACTTGAATGTTCCCCCGCTTT***GACGCTCATACAGGGTCAAAGGGGTAAGG**
N. crassa mtA-3

FIGURE 5.2.3b) The sequence at the junction of mtA/mt**a** at *S. fimicola* compared to the corresponding sequences in *N. crassa* mtA and mt**a**. The nucleotides in bold show the nucleotides relevant to *S. fimicola*, those nucleotides in italics do not have corresponding sequence in *S. fimicola.*

Ideally the sequences of mt**a** and mtA in the ancestor of *S. fimicola* would be examined to get a clearer idea of how the mtA and mt**a** idiomorphs recombined at this point. A small region of limited homology, for example, might facilitate an unequal crossing over event which would bring the idiomorphs together. A comparison of *S. fimicola* with its immediate ancestor is obviously not possible so the mt**a** and mtA-3 sequences of *N. crassa* were used instead. As the idiomorphs are under evolutionary constraints it is not impossible that the *N. crassa* sequences may give insight into the recombination event. However there are no obvious reasons for recombination taking place at the junction site, such as homologous sequences which could have paired to give unequal crossing over. The fact that the *S. fimicola* mtA-3 and mt**a**-1 ORFs are transcribed in the same direction as each other, which is the case when the two idiomorphs are observed separately in heterothallic species, suggests that it could have been an unequal crossing over event that recombined the idiomorphs in this fashion. At the end of *S. fimicola* mt**a**-1 sequences corresponding to centromere distal sequences were found (see chapter 6) and although centromere proximal sequences were not obtained for *S. fimicola*, an *Nsi* I site is present downstream of mtA-1 indicating that centromere proximal sequences are possibly present. Again this is evidence that the mtA and mt**a** chromosomes in the ancestor of *S. fimicola* could have aligned in such a manner to produce unequal crossing over and so

maintain the idiomorphic flanking sequences in their correct positions relative to the centromere.

MtA-3 in *S. fimicola* runs into mt**a** 5'-3'. The sequencing data for mt**a** was examined for stop codons and an in frame TGA stop codon was found 77 nucleotides downstream of the junction site (data not shown). It is possible that MtA-3 is a non-functional protein and not required for post-fertilization events in *S. fimicola.*

Chapter 6

Evolution Within the Sordariaceae

6.1 Introduction

The mating type genes of the Sordariaceae contain several distinct regions which are of interest when drawing up comparisons in an evolutionary context. In this chapter the DNA sequence data detailed in chapters 3, 4 and 5 are compared. Section 6.2.1 details the alignments of the DNA sequences and amino acid sequences of the mtA-1 and mt**a**-1 *Sordaria* genes sequenced in this project and the mtA-1 and mt**a**-1 *Neurospora* genes available from the database. From the mtA-1 alignment a phylogenetic analysis was carried out to establish the evolutionary relationships between the species examined. Interesting features from available sequence data upstream and downstream of the mating type genes are commented on in section 6.2.2. The common regions of *S. equina* and *S. sclerogenia* were compared to the common regions of *Neurospora* species and the variable regions compared to see if any sequence similarity exists between the *Neurospora* and *Sordaria* species.

6.2 Results

6.2.1 Comparisons Between mtA-1 and mta-1 genes of *Sordaria* and *Neurospora* Species.

In figure 6.2.1a) the DNA sequences of the mtA-1 genes from *N. sitophila, N. crassa, N. intermedia, N. tetrasperma, N. discreta, N. africana, S. fimicola, S. equina* and *S. sclerogenia* are aligned using CLUSTALW. One can see that the position of the intron in the mtA-1 gene in all the species is conserved. Seven out of the nine species have a 5' splice sequence of GTAAGT. *N. crassa* and *N. sitophila* both have the 5' splice sequence of GTGAGT. The 3' splice sequence is CAG in all species except for *N. intermedia* which has TAG. The alignment of the species between positions 125 and145 is of interest as all the species appear to have a deletion in this region when compared to *S. fimicola*. (or *S. fimicola* has an insert in relation to the other species). A similar feature is seen between positions 390 and 407, *S. fimicola* here has an extra 18 nucleotides compared to the other species. Between positions 620 and 625 all the *Neurospora* species and *S. fimicola* have a 6 nucleotide deletion when compared to the remaining *Sordaria* species (or *S. equina* and *S. sclerogenia* have a 6 nucleotide insert compared to the other species). These inserts/deletions are all in multiples of three so do not disrupt the open reading frame.

```
1 N.crassa        ATGTCGGGTGTCGAT CAAATCGTCAAGACG TTCGCCGACCTCGCT GAGGACGACCGTGAA GCGGCAATGAGAGCT TTCTCAAGGATGATG
2 N.sitophila     ATGTCGGGAGTCGAT CAAATCGTCAAGACG TTCGCCGACCTCGCT GAGGACGACCGTGAA GCGGCAATGAGAGCT TTCTCAACGATGATG
3 N.intermedia    ATGTCGGGTGTCGAT CAAATCGTCAAGACG TTCGCCGACCTCGCT GAGGACGACCGTGAA GCGGCAATGAGAGCT TTCTCAACGATGATG
4 N.tetrasperma   ATGTCGGGTGTCGAT CAAATCGTCAAGACG TTCGCCGACCTCGCT GAGGACGACCGTGAA GCGGCAATGAGAGCT TTCTCAACGATGATG
5 N.discreta      ATGTCGGGCGTCGAC CAAATTGTCAAGACG TTCGCCGACCTCGCT GAGGACGACCGCGAA G---CAATGAGAGCT TTCTCAACGATGATG
6 N.africana      ATGTCGTGCGTCGAT CAAATCGTCAAGACG TTCGCCGACCTCACT GAGGGTGATCGTGAA GCGGCAATGAGAGCT TTCTCAATGATGATG
7 S.fimicola      ATGTCCAGCGTCGAT CAAATCGTCAAGACG TTCGCCAACCTCCCT GAGGGCGAGCGCAAC GCAGCAGTCAATGCT ATCTTAGCCATGATG
8 S.sclerogenia   ATGTCAGGCGTCGAT CAAATCGTCAAAACG TTCGCCAATCTCGGT GAGGGCGATCGTGAA GCGGCCATGAAAGCT TTCTTAGCGATGATG
9 S.equina        ATGTCAGGCGTCGAT CAAATCGTCAAAAAG TTCGCCAATCTCGGT GAGGGTGATCGTGAA GCGGCCATGAAAGCT TTCTTAGCGATGATG

                  91          105 106         120 121
1 N.crassa        CGTAGAGGT------ --------------- ACCGAACCTGTTCGC CGAATCCCCGCGGCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
2 N.sitophila     CGT------------ --------------- ACCGAACCTGTTCGC CGAATCCCCGCGGCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
3 N.intermedia    CGT------------ --------------- ACCGAACCTGTTCGC CGAATCCCCGCGGCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
4 N.tetrasperma   CGT------------ --------------- ACCGAACCTGTTCGC CGAATCCCCGCGGCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
5 N.discreta      CGT------------ --------------- ACCGAACCTGTTCGC CAAATCCCCGCGACA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
6 N.africana      CGC------------ --------------- ACCGAACCTGTTCGC CAAACCCCCGCGGCA AAGAAGAAGGTCAAC GGCTTCATGAGTTTC
7 S.fimicola      CCCCCCGGCCCTGGT CCTGTTCGCCAAATC CCCGAACCTGTTCCA CAAGCCCCCGCGCCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
8 S.sclerogenia   CCCGTGAGCAACGAA ACT---------GTC GCCGAACCTGTCCGC AAAGCCCCTGCCGCG AAGAAGAAGGTCAAC GGCTTCATGGGTTTC
9 S.equina        CCCGTGAGCAACGAA CCT---------GTC GCTGAACCTGTCCGC AAAGCCCCCACCGCA AAGAAGAAGGTCAAC GGCTTCATGGGTTTC

1 N.crassa        AGATgtgagtCAAAT CTGAATCAACATTGT CGTT-GATCCATGGC TGATTGCTCTTC-AT TTcagCGTACTATTC CCCGCTCTTCTCTCA
2 N.sitophila     AGATgtgagtCAAAT CTGAATCAACATTGT CGTT-GATCCATGGC TGATTGCTCTTC-AT TTcagCGTACTATTC CCCGCTCTTCTCTCA
3 N.intermedia    AGATgtaagtCAAAT CTGAATCAACATTGT CGTT-CATCCATGGC TAATTGCTCTTC-AA TTtagCGTACTATTC CCCGCTCTTCTCTCA
4 N.tetrasperma   AGATgtaagtCAAAT CTGAATCAACATTGT CGTTTGATCCATGGC TAATTGCTCTTC-AT TTcagCGTACTATTC CCCGCTCTTCTCTCA
5 N.discreta      AGATgtaagtTAAAT CTGAATCAATCTTGT CGAT-AATCCATG-C TGACTGCTCTTC-AT TTcagCGTACTATTC CCCGCTCTTCTCTCA
6 N.africana      AGATgtaagtCAAAT CTGGATCAATCTTGT TGAA-AATCCATT-C TAATTGCCTTTT-AT TTcagCGTACTATTC CCCGCTCTTCTCTCA
7 S.fimicola      AGATgtaagtCAAAT CTGAATTAATCTTGA CGAC-GATCCATA-C TGATTGCCTTTCTAT TTcagCGTACTATTC CTCGCTCTTCTCTCA
8 S.sclerogenia   AGATgtaagtCAGAT CTGAGTCAATCTCGT GGAC-AGTCCATA-C TAATTGTTTTTT--C TTcagCGTACTATTC CCCGCTCTTCTCTTA
9 S.equina        AGATgtaagtCAGAT CTGAGTCAATCTTGT CGAC-AGTCTATG-C TAATTGTTTTTC--C TTcagCGAACTATTC CCCGCTCTTCTCTTA
```

```
1 N.crassa       GCTCCCGCAAAAGGA GAGATCGCCCTTCAT GACTATTCTCTGGCA GCATGATCCCTTCCA CAATGAGTGGGATTT CATGTGCTCGGTGTA
2 N.sitophila    GCTCCCGCAAAAGGA GAGATCGCCCTTCAT GACCATTCTCTGGCA GCATGATCCCTTCCA CAATGAGTGGGATTT CATGTGCTCGGTGTA
3 N.intermedia   GCTCCCGCAAAAGGA GAGATCGCCCTTCAT GACCATTCTCTGGCA GCATGATCCCTTTCA CAATGAGTGGGATTT CATGTGCTCGGTGTA
4 N.tetrasperma  GCTCCCGCAAAAGGA GAGATCGCCCTTCAT GACCATTCTCTGGCA GCATGATCCCTTCCA CAATGAATGGGATTT CATGTGCTCGGTGTA
5 N.discreta     GCTCCCGCAAAAGGA GAGATCGCCCTTCAT GACCATTCTTTGGCA GCACGATCCCTTCCA CAACGAATGGGATTT CATGTGCTCGGTGTA
6 N.africana     GCTCCCGCAGAAGGA GAGATCACCCTTCAT GACCATTCTCTGGCA GCACGATCCCTTCCA CAACGAATGGAATTT CATGTGCTCGGTGTA
7 S.fimicola     GTTTCCTCAGAAGGC GCGATCGCCCTTCAT GACCATCCTCTGGCA GCACGATCCCTTTCA CAACGAATGGGATTT CATGTGCTCGGTGTA
8 S.sclerogenia  CCTTCCACAAAAGAT GAGGTCGCCTTTCAT GACCATTCTCTGGCA GTACGATCCCTACCA CACCGAATGGGATTT CATGTGTTCGGTTTA
9 S.equina       CCTTCCGCAAAAGAT GAGGTCGCCTTTCAT GACCATTCTCTGGCA GTACGACCCCTACCA CAACGAATGGGATTT CATGTGTTCGGTGTA

                                                  390 391         405 406
1 N.crassa       TTCGTCAATCCGGAC CTACCTTGAGCAGGA G-------------- ----AAGGTTACTCT GCAACTCTGGATTCA CTATGCTGTCGGCCA
2 N.sitophila    TTCGTCAATCCGGAC CTACCTTGAGCAGGA G-------------- ----AAGGTTACTCT GCAACTCTGGATTCA CTATGCTGTCGGCCA
3 N.intermedia   TTCGTCAATCCGGAC CTACCTTGAGCAGGA G-------------- ----AAGGTTACTCT GCAACTCTGGATTCA CTATGCTGTCGGCCA
4 N.tetrasperma  TTCGTCAATCCGGAC CTACCTTGAGCAGGA G-------------- ----AAGGTTACTCT GCAACTCTGGATTCA CTATGCTGTTCGCCA
5 N.discreta     TTCGTCAATCCGTAC CTATCTTGAGCAGGA G-------------- ----AAGGTTACTCT GCAACTTTGGATTCA CTATGCTGTCGGCCA
6 N.africana     TTCGTCGATCCGCAC CTACCTTGAGCAGGA G-------------- ----AAAGTTACCCT GCAACTCTGGATTCA CTATCGTGTCCGCCA
7 S.fimicola     TTCGTCAATCCGCAA CTACCTCGAGCAGTT GAACGCGCAGCGGGA GAAGAAGATTACCCT GCAATACTGGCTTCA CTTTGCTGTCCCCGT
8 S.sclerogenia  TTCCTCAATCCGCAC CGACCTGGAGGAGGA G-------------- ----AAGGTTACACT GCAGCTCTGGATTCA CTATGCTATCGGCCA
9 S.equina       TTCTTCAATCCGCAC CGACCTGGAGGAGCA G-------------- ----AATGTTACACT GCAGCTCTGGATTCA CTATGCTATCGGCCA


1 N.crassa       TCTGGGAGTGATTAT CCGCGACAACTACAT GGCATCCTTTGGCTG GAACCTCGTCCGTTT TCCCAACGGCACTCA CGACCTCGAGCGCAC
2 N.sitophila    TCTGGGAGTGATTAT CCGCGACAACTACAT GGCATCCTTTGGCTG GAACCTCGTCCGTTT TCCCAACGGCACTCA CGACCTCGAGCGCAC
3 N.intermedia   TCTGGGAGTGATTAT CCGCGATAACTACAT GGCATCGTTTGGTTG GAACCTCGTCCGTTT TCCCAACGGCACTCA CGACCTCGAGCGCAC
4 N.tetrasperma  TCTGGGAGTGATTAT CCGCGACAACTACAT GGCATCGTTTGGCTG GAACCTCGTCCGTTT TCCCAACGGCACTCA CGACCTCGAGCGCAC
5 N.discreta     TCTGGGAGTGATTAT CCGCGACAACTACAT GGCATCGTTTGGCTG GAACCTCGTCCATCT GCCCAACGGCACGCA CGACCTCGAGCGCAC
6 N.africana     TCTGGGAGTGATTAT CCGCGACAACTACAT GGCATCGTTTGGCTG GAACCTCGTCCAGCT GCCCAACGGCACTCA CGACCTCGAGCGCAC
7 S.fimicola     CATGGGAGTGCTTGG TCGCGAAAACTACTT GCCCACGCTTGGCTG GGACCTCGTCACGAT GCCCAACGGCACTAT CGACCTTATGCGCAT
8 S.sclerogenia  GATGGGATTGATTGA CCGCGACCACTACAT GGCATCGTTTGGCTG GCGCCTCGGTCAGAC TCGCAACGGCACTAC CGACCTTTTTCGCAC
9 S.equina       GATGGGATTGATTGA CCGCGACCACTACAT GGCATCGTTTGGCTG GCGCCTCGGTCAGAC TCGCAACGGCACTAC CGACCTTTTTCGCAC
```

```
                                                                                                616            630
1 N.crassa      GGCTCTTCCTTTGGT TCAGCACAATCTCCA GCCCATGAACGGCTT ATGCCTGCTCACCAA GTGCCTCGAGAGCGG ATTGCCT------CT
2 N.sitophila   GGCTCTTCCTTTGGT TCAGCACAATCTCCA GCCCATGAACGGCCT ATGCCTGCTCACCAA GTGCCTCGAGAGCGG ATTGCCT------CT
3 N.intermedia  GGCTCTTCCTTTGGT TCAGCCCAATCTCCA GCCCATGAACGGCTT ATGCCTGCTCACCAA GTGCCTCGAGAGCGG ATTGCCT------CT
4 N.tetrasperma GGCTCTTCCTTTGGT TCAGCACAATCTCCA GCCCATGAACGGCTT ATGCCTGCTCACCAA GTGCCTCGAGAGCGG ATTGCCT------CT
5 N.discreta    CGCTCTTCCCTTGGT TAGGCACAATCTCCA GCCCATGAACGGCCT ATGCCTGCTCACTAA GTGCCTTGAGAGCGG ATTACCT------CT
6 N.africana    CGCTCTTCCTTTGGT TCAGCATAACCTCCA GCCCATGAACGGCCT ATGCCTGTTCACCAA GTGCCTCGAGAGCGG ATTGCCT------CT
7 S.fimicola    CGCTATGCCTTTGTT TAGAAAGAACCTCCA GCCCATGGACGGCCT ATGCCTGTTCACCAA GTGTCAGGAGGGCGG ATTGCAA------GT
8 S.sclerogenia TGCGATTCCGATGGT TAGGCGCAAGCTTCA GCCCATGAACGGCCT TTGCCTGCTCATTAA GTGTCTCCAGAGCGG ATTGAACAAGCATCT
9 S.equina      TGCGATTCCGATGGT TAGGCGCAACCTTCA GCCCATGAACGGCCT TTGCCTGCTCATTAA GTGTCTCGAGAGCGG ATTGAGCAAGCATCT

1 N.crassa      TGCCAATCCTCACTC TGTCATCGCCAAGCT TTCAGATCCTAGCTA CGACATGATCTGGTT CAACAAGCGTCCTCA CCGTCAGCAGGGACA
2 N.sitophila   TGCCAATCCTCATTC TGTCATCGCCAAGCT TTCAGATCCTAGCTA CGACATGATCTGGTT CAACAAGCGTCCTCA CCGTCAGCAGGGACA
3 N.intermedia  TGCCAATCCTCACTC TGTCATCGCCAAGCT TTCAGATCCTAGCTT CGACATGATCTGGTT CAACAAGCGTCCTCA CCGTCAGCAGGGACA
4 N.tetrasperma TGCCAATCCTCACTC TGTCATCGCCAAGCT TTCAGATCCTAGCTA TGACATGATCTGGTT CAACAAGCGTCCTCA CCGTCAGCAGGGACA
5 N.discreta    TGCCAATCCTCACTC TGTCATCGCCAAGCT TTCAGATCCTAGCTA CGACATGATCTGGTT CAACAAGCGTCCTTA CAGTCAGCAGAGACA
6 N.africana    TGCCAATCCTCACCC TGTCATCGCCAAGCT TTCAGATCCTAGCTA CGACATGATCTGGTT CAACAAGCGTCCTCA CCGTCAGCAGGGACA
7 S.fimicola    CGACAACCAGCACTT CGTCATTGCCAAGCT TTCAGATCCTAGCCA CGACATGATCTGGTT CAACAAGCGCCCTCA CTATCAGCAGAGACA
8 S.sclerogenia TACCAATCCTCATCC TGTTATTGCCAAGCT GCAAGATCCTAGCTT CGACATGATTTGGAT CAACAAGCCTTCTCA CCATCAGCAGGGACA
9 S.equina      TACCAATCCCCATCC TATTATCGCCAAGCT GCAAGATCCTAGCTT CGACATGATCTGGAT CAACAAGCCTCCTCA CCATCAGCAGGGACA

1 N.crassa      CGCCGTTCAAACTGA TGAATCTGAAGTTGG AGTTTCGGCGATGTT CCCTCGCAATCACAC GGTCGCTGCAGAGGT AGATGGCATCATCAA
2 N.sitophila   CGCCGTTCAAACTGA TGGATCTGAAGTTGG AGTTTCGGCGATGTT CCCTCGCAATCACAC GGTCGCTGCAGAGGT AGATGGCATCATCAC
3 N.intermedia  CGCCGGTCAAACTGA TGAATCTGAAGTTGG AGTTTCGGCGATGTT CCCTCGCAATCACAC GGTCGCTACAGAGGT AGATGGCATCATCAA
4 N.tetrasperma CGTCGGTCAAACTGA TGAATCTGAGGTTGG AGTTTCGGCGATGTT CCCTCGCAATCACAC GGTCGCTGCCGAGGT AGATGGCATCATCAA
5 N.discreta    AGTCGGCCAAACTGA CGATTCTGAACTCGA AGTGTCGGCGATGTT CCCTCACAATTACGC AGTCGCCGCAGAGGC AGATGGTATCGCCAA
6 N.africana    CGCCGGCCAAACTTA CAATTCTGAACTTGG AGTGTCGGCGCTCTT CCCTTGCAATCACGC AGTCGCTGCAGCGGT CGATGGCATCACCGA
7 S.fimicola    CGCCGCCCAAACCGA CAGTTCTGAACTCGG TGTGTCGGCGCTCTT CCCTCGCAATCACGC AGTTGCTGCAGAGGC AGATGGCGTCGCCAC
8 S.sclerogenia CACCGATCAAGCTGA CAATTCTGAACTCAG AATGCCGTCGATCTT CCCTAGCAATCACGC AGTCGCTGCGGAGGT AGATGGCATCGCCAA
9 S.equina      CACCGACCAAGCTGA CAATTCCGAACTCGG AATGCCGTCGCTCTT CCCTGGAAATCATGC AGTCGCTGCGGAGGT AGATGGCATCGCCAA
```

```
1 N.crassa       TCTTCCTCTCTCCCA TTGGATTCAGCAGGG AGAATTCGGTACCGA GTCTGGATACTCAGC TCAGTTTGAGACCTT GTTGGATTCAATTCT
2 N.sitophila    ---TCCTCTCTCCCA TTGGATTCAGCAGGG AGAATTCGGTACCGA GTCTGGATACTCAGC TCAGTTTGAGACCTT GTTGGATTCAATTCT
3 N.intermedia   TCTTCCTCTCTCCCA TTGGATTCAGCAGGG AGAATTCGGTACCGA GTCTGGATACTCAGC TCAGTTTGAGACATT GTTGGATTCAATCCT
4 N.tetrasperma  TCTTCCTCTCTCCCA TTGGATTCAGCAGGG AGAATTCGGTACCGA GTCTGGATACTCAGC TCAGTTTGAGACCTT GTTGGATTCAATCCT
5 N.discreta     TCTTCCTCTCTCCCA TTGGATTCAGCAGGG AGATTTCGGTACTGA CCCCGGATACTCAGC TCAATTTGAGACTTT GTTGGATTCTATTCT
6 N.africana     CCTTCCTCTCTCCCA TTGGCTTCAGCAGGG AGATTTCGGCACCGA GGCCGGATTCTCACC TCAGTTTGAGACCTT GTTGGATTCGATCCT
7 S.fimicola     TGTTCAACTCCCTCA TTGGATGCAGCAGGG AGATTTCGGCACCGA GTCCGGATACTCACC TCAGTTTGAGACCTT GTTGGGTTCCATACT
8 S.sclerogenia  CCTTCCTCTCTCACA TTGGACTCAGCAGGG GGATTTCGGCAAGGA GCCTGGATTCTCGAC CCAGTTTGATACCAT GTTGGATTCACTTCT
9 S.equina       CCTTCCTCTCTCACA TCGGACTCAGCAGGG AGATTTCGGCACCGA GCCTGGATTCTCGAC TCAGTTTGATACCAT GTTGGATTCAATTCT

1 N.crassa       CGAGAATGGACACGC CTCCAGCAATGACCC TTACAACATGGCTCT GGCTATCGATGTTCC CATGATGGGTTAG
2 N.sitophila    CGAGAATGGACACGC CTCCAGCAATGACCC TTACAACATGGCTCT GGCTATCGATGTTCC CATGATGGGTTAG
3 N.intermedia   CGAGAATGGACACGC CTCCAGCAATGACCC TTACAACATGGCTCT GGCTATCGATGTTCC CATGATGGGTTAG
4 N.tetrasperma  CGAGAATGGACACGC CTCCAGCAATGACCC CTACAACATGGCTCT GGCTATCGATGTTCC CATGATGG-TTAG
5 N.discreta     TGAGGATGGACACGC CTCCAGCAATGACCC CTACAACATGGCTCT GGCTATGGATGTTCC CATGATGGGTTAG
6 N.africana     TGAGAATGGAAACGC CTCTATCAATGACCC CTACAATATGGCTCT TGGTATGGGTGTTCC CATGATGGGTTAG
7 S.fimicola     TGAGAATGGAAACGC CACCAGTAATGATTC CTACAACATGGCTCT GGCTATGGATGTTCC TATGATGGGTTAG
8 S.sclerogenia  TGAGAATGGAAACGA CACCAGCAATCATCA CTACAACATGGCTCT GGCTATGGATCTTCC CATGATAGGTTAG
9 S.equina       CGAGAATGGAAACCA TCCCAGCAATAGTCA CTACAACATGTCTCT GGCTATGGATCTTCC TATGACGGGTTAG
```


FIGURE 6.2.1a)

Multiple alignment of the mtA-1 genes from *N. crassa, N. sitophila, N. intermedia, N. tetrasperma, N. discreta, S. fimicola, N. africana, S. sclerogenia* and *S. equina* using CLUSTALW. The 5' and 3' intron splice sites are shown in lower case. The intron is shown in italics and underlined.

```
                                                  31          45
1 N.sitophila    MSGVDQIVKTFADLA EDDREAAMRAFSTMM R---------TEPVR RIPAAKKKVNGFMGF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
2 N.crassa       MSGVDQIVKTFADLA EDDREAAMRAFSRMM RRG-------TEPVR RIPAAKKKVNGFMGF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
3 N.intermedia   MSGVDQIVKTFADLA EDDREAAMRAFSTMM R---------TEPVR RIPAAKKKVNGFMGF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
4 N.tetrasperma MSGVDQIVKTFADLA EDDREAAMRAFSTMM R---------TEPVR RIPAAKKKVNGFMGF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
5 N.discreta     MSGVDQIVKTFADLA EDDR-EAMRAFSTMM R---------TEPVR QIPATKKKVNGFMGF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
6 N.africana     MSCVDQIVKTFADLT EGDREAAMRAFSMMM R---------TEPVR QTPAAKKKVNGFMSF RSYYSPLFSQLPQKE RSPFMTILWQHDPFH
7 S.fimicola     MSSVDQIVKTFANLP EGERNAAVNAILAMM PPGPGPVRQIPEPVP QAPAPKKKVNGFMGF RSYYSSLFSQFPQKA RSPFMTILWQHDPFH
8 S.equina       MSGVDQIVKKFANLG EGDREAAMKAFLAMM PVSNEP---VAEPVR KAPTAKKKVNGFMGF RSNYSPLFSYLPQKM RSPFMTILWQYDPYH
9 S.sclerogenia MSGVDQIVKTFANLG EGDREAAMKAFLAMM PVSNET---VAEPVR KAPAAKKKVNGFMGF RSYYSPLFSYLPQKM RSPFMTILWQYDPYH

                                 106         120
1 N.sitophila    NEWDFMCSVYSSIRT YLEQE------KVTL QLWIHYAVGHLGVII RDNYMASFGWNLVRF PNGTHDLERTALPLV QHNLQPMNGLCLLTK
2 N.crassa       NEWDFMCSVYSSIRT YLEQE------KVTL QLWIHYAVGHLGVII RDNYMASFGWNLVRF PNGTHDLERTALPLV QHNLQPMNGLCLLTK
3 N.intermedia   NEWDFMCSVYSSIRT YLEQE------KVTL QLWIHYAVGHLGVII RDNYMASFGWNLVRF PNGTHDLERTALPLV QPNLQPMNGLCLLTK
4 N.tetrasperma NEWDFMCSVYSSIRT YLEQE------KVTL QLWIHYAVRHLGVII RDNYMASFGWNLVRF PNGTHDLERTALPLV QHNLQPMNGLCLLTK
5 N.discreta     NEWDFMCSVYSSIRT YLEQE------KVTL QLWIHYAVGHLGVII RDNYMASFGWNLVHL PNGTHDLERTALPLV RHNLQPMNGLCLLTK
6 N.africana     NEWNFMCSVYSSIRT YLEQE------KVTL QLWIHYRVRHLGVII RDNYMASFGWNLVQL PNGTHDLERTALPLV QHNLQPMNGLCLFTK
7 S.fimicola     NEWDFMCSVYSSIRN YLEQLNAQREKKITL QYWLHFAVPVMGVLG RENYLPTLGWDLVTM PNGTIDLMRIAMPLF RKNLQPMDGLCLFTK
8 S.equina       NEWDFMCSVYSSIRT DLEEQ------NVTL QLWIHYAIGQMGLID RDHYMASFGWRLGQT RNGTTDLFRTAIPMV RRNLQPMNGLCLLIK
9 S.sclerogenia TEWDFMCSVYSSIRT DLEEE------KVTL QLWIHYAIGQMGLID RDHYMASFGWRLGQT RNGTTDLFRTAIPMV RRKLQPMNGLCLLIK

1 N.sitophila    CLESGLP--LANPHS VIAKLSDPSYDMIWF NKRPHRQQGHAVQTD GSEVGVSAMFPRNHT VAAEVDGIITP-LSH WIQQGEFGTESGYSA
2 N.crassa       CLESGLP--LANPHS VIAKLSDPSYDMIWF NKRPHRQQGHAVQTD ESEVGVSAMFPRNHT VAAEVDGIINLPLSH WIQQGEFGTESGYSA
3 N.intermedia   CLESGLP--LANPHS VIAKLSDPSFDMIWF NKRPHRQQGHAGQTD ESEVGVSAMFPRNHT VATEVDGIINLPLSH WIQQGEFGTESGYSA
4 N.tetrasperma CLESGLP--LANPHS VIAKLSDPSYDMIWF NKRPHRQQGHVGQTD ESEVGVSAMFPRNHT VAAEVDGIINLPLSH WIQQGEFGTESGYSA
5 N.discreta     CLESGLP--LANPHS VIAKLSDPSYDMIWF NKRPYSQQRQVGQTD DSELEVSAMFPHNYA VAAEADGIANLPLSH WIQQGDFGTDPGYSA
6 N.africana     CLESGLP--LANPHP VIAKLSDPSYDMIWF NKRPHRQQGHAGQTY NSELGVSALFPCNHA VAAAVDGITDLPLSH WLQQGDFGTEAGFSP
7 S.fimicola     CQEGGLQ--VDNQHF VIAKLSDPSHDMIWF NKRPHYQQRHAAQTD SSELGVSALFPRNHA VAAEADGVATVQLPH WMQQGDFGTESGYSP
8 S.equina       CLESGLSKHLTNPHP IIAKLQDPSFDMIWI NKPPHHQQGHTDQAD NSELGMPSLFPGNHA VAAEVDGIANLPLSH RTQQGDFGTEPGFST
9 S.sclerogenia CLQSGLNKHLTNPHP VIAKLQDPSFDMIWI NKPSHHQQGHTDQAD NSELRMPSIFPSNHA VAAEVDGIANLPLSH WTQQGDFGKEPGFST
```

```
1 N.sitophila    QFETLLDSILENGHA SSNDPYNMALAIDVP MMG
2 N.crassa      QFETLLDSILENGHA SSNDPYNMALAIDVP MMG
3 N.intermedia  QFETLLDSILENGHA SSNDPYNMALAIDVP MMG
4 N.tetrasperma QFETLLDSILENGHA SSNDPYNMALAIDVP MMV
5 N.discreta    QFETLLDSILEDGHA SSNDPYNMALAMDVP MMG
6 N.africana    QFETLLDSILENGNA SINDPYNMALGMGVP MMG
7 S.fimicola    QFETLLGSILENGNA TSNDSYNMALAMDVP MMG
8 S.equina      QFDTMLDSILENGNH PSNSHYNMSLAMDLP MTG
9 S.sclerogenia QFDTMLDSLLENGND TSNHHYNMALAMDLP MIG
```

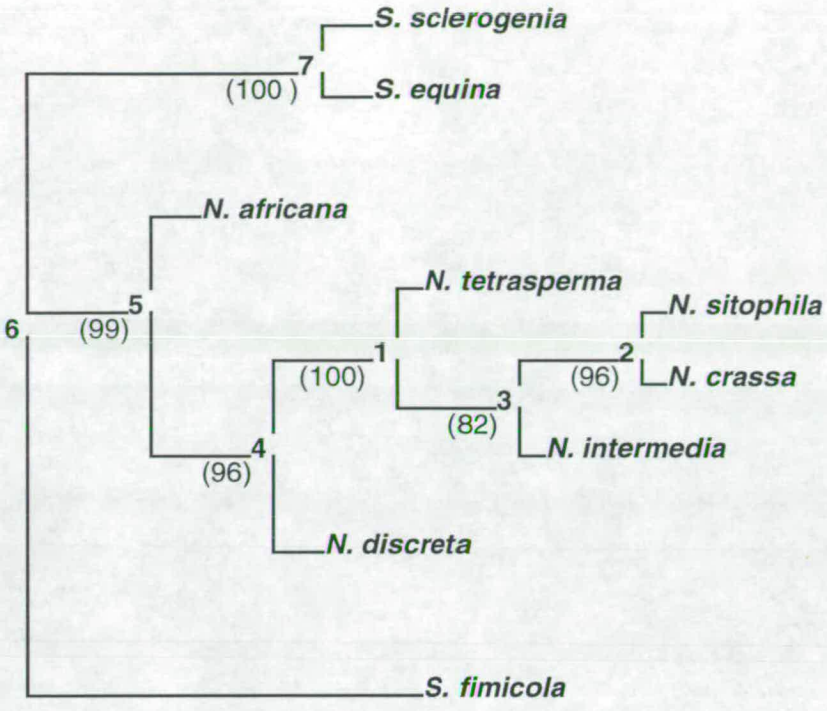
FIGURE 6.2.1b)

Multiple alignments of the amino acid sequences of the mtA-1 genes from the species examined in figure 6.2.1a). The putitive DNA binding domain, the α domain , is shown in italics and underlined.

Figure 6.2.1b) shows the multiple alignment of the amino acid sequences of the mtA-1 genes from the species compared in figure 6.2.1a). Again the multiple alignments were produced using the CLUSTALW programme. As expected from the DNA sequence *S. fimicola* has extra amino acids between positions 32-40 and 111-116 when compared to the other species. The extra amino acids between position 111-116 in *S. fimicola* interrupts a region shown by Saupe *et al* (1996) to be involved in vegetative incompatibility function. This is of interest because *S. fimicola* contains both mating types in the same nucleus so some change to the region involved in vegetative incompatibility could have made this scenario possible. Amino acid changes between the species at the α domain are conservative and the α domain is well conserved.

Codon usage in the *Sordaria* mtA-1 genes sequenced was typical of codon usage in *Neurospora*.

The multiple alignments of the mtA-1 DNA sequences were used to build a phylogeny to see how the various *Sordaria* and *Neurospora* species are related (for methodology see section 2.2.4.2). This phylogeny is shown in figure 6.2.1c). A phylogeny constructed using the amino acid alignments of the species produced a tree with the same topology as that shown in figure 6.2.1c)(not shown).



Between	And	Length
6	fimicola	0.14778
6	7	0.12241
7	sclerogenia	0.02974
7	equina	0.02772
6	5	0.02807
5	africana	0.04622
5	4	0.01895
4	1	0.03474
1	tetrasperma	0.00656
1	3	0.00243
3	2	0.00556
2	sitophila	0.00549
2	crassa	0.00224
3	intermedia	0.01128
4	discreta	0.04217

FIGURE 6.2.1c)

An unrooted tree demonstrating the evolutionary relationships between *Sordaria* and *Neurospora* species. The branch lengths, corresponding to the rate of nucleotide substitutions per base, are shown in the table underneath. The numbers in brackets indicate the number of times the group consisting of the species which are to the right of the forks occurred among the trees, out of 100 trees, during bootstrapping analysis. Values over 70 provide good support that the clade is accurate.

The phylogeny shown in figure 6.2.1c) shows that *S. fimicola* is only distantly related to the other *Sordaria* species used in the analysis, *S. sclerogenia* and *S. equina* , which appear to be closely related. Both *S. fimicola* and *S. equina* are homothallic species and one can see from this phylogeny that these two species appear to have evolved separately. Metzenberg and Glass (1990) proposed an evolutionary scenario where a species like *S. equina* might evolve from a species like *S. fimicola* by loss of the mt**a** idiomorph. This doesn't appear to be the case judging from the phylogeny shown in figure 6.2.1c). However one must be cautious interpreting this tree as a limited amount of data was used in its production. Obviously more sequence data from other *Neurospora* and *Sordaria* species would make the tree more accurate. Other problems with basing a phylogeny on mating type idiomorph data will be discussed later.

Figure 6.2.1d) shows the *N. crassa* mt**a**-1 DNA sequence aligned with the *S. fimicola* mt**a**-1 DNA sequence. *N. crassa* mta-1 is the only mt**a**-1 *Neurospora* gene available from the database.

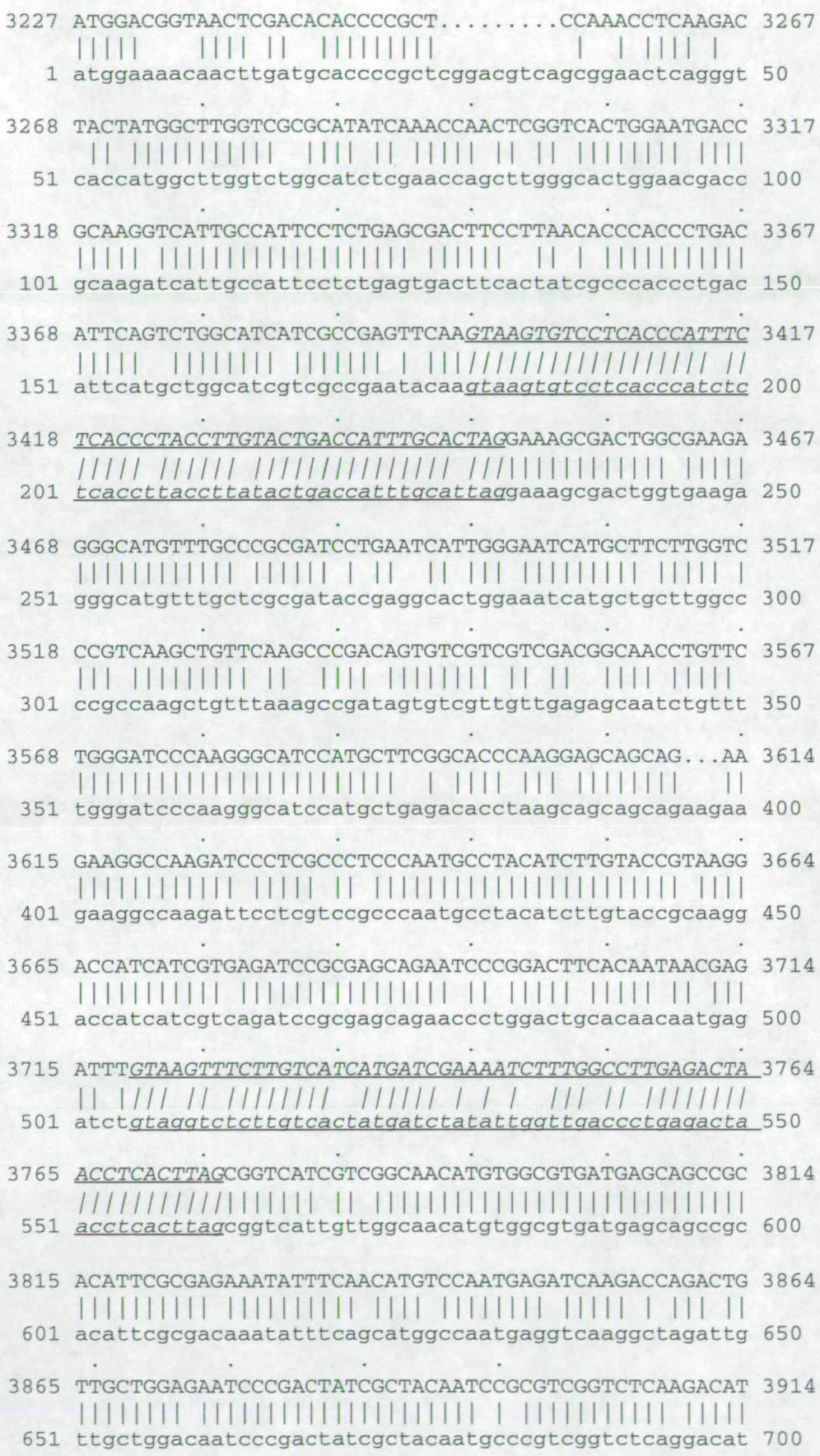

3227 ATGGACGGTAACTCGACACACCCCGCT.....CCAAACCTCAAGAC 3267
1 atggaaaacaacttgatgcaccccgctcggacgtcagcggaactcagggt 50
3268 TACTATGGCTTGGTCGCGCATATCAAACCAACTCGGTCACTGGAATGACC 3317
51 caccatggcttggtctggcatctcgaaccagcttgggcactggaacgacc 100
3318 GCAAGGTCATTGCCATTCCTCTGAGCGACTTCCTTAACACCCACCCTGAC 3367
101 gcaagatcattgccattcctctgagtgacttcactatcgcccaccctgac 150
3368 ATTCAGTCTGGCATCATCGCCGAGTTCAAGTAAGTGTCCTCACCCATTTC 3417
151 attcatgctggcatcgtcgccgaatacaagtaagtgtcctcacccatctc 200
3418 TCACCCTACCTTGTACTGACCATTTGCACTAGGAAAGCGACTGGCGAAGA 3467
201 tcaccttaccttatactgaccatttgcattaggaaagcgactggtgaaga 250
3468 GGGCATGTTTGCCCGCGATCCTGAATCATTGGGAATCATGCTTCTTGGTC 3517
251 gggcatgtttgctcgcgataccgaggcactggaaatcatgctgcttggcc 300
3518 CCGTCAAGCTGTTCAAGCCCGACAGTGTCGTCGTCGACGGCAACCTGTTC 3567
301 ccgccaagctgtttaaagccgatagtgtcgttgttgagagcaatctgttt 350
3568 TGGGATCCCAAGGGCATCCATGCTTCGGCACCCAAGGAGCAGCAG...AA 3614
351 tgggatcccaagggcatccatgctgagacacctaagcagcagcagaagaa 400
3615 GAAGGCCAAGATCCCTCGCCCTCCCAATGCCTACATCTTGTACCGTAAGG 3664
401 gaaggccaagattcctcgtccgcccaatgcctacatcttgtaccgcaagg 450
3665 ACCATCATCGTGAGATCCGCGAGCAGAATCCCGGACTTCACAATAACGAG 3714
451 accatcatcgtcagatccgcgagcagaaccctggactgcacaacaatgag 500
3715 ATTTGTAAGTTTCTTGTCATCATGATCGAAAATCTTTGGCCTTGAGACTA 3764
501 atctgtaggtctcttgtcactatgatctatattggttgaccctgagacta 550
3765 ACCTCACTTAGCGGTCATCGTCGGCAACATGTGGCGTGATGAGCAGCCGC 3814
551 acctcacttagcggtcattgttggcaacatgtggcgtgatgagcagccgc 600
3815 ACATTCGCGAGAAATATTTCAACATGTCCAATGAGATCAAGACCAGACTG 3864
601 acattcgcgacaaatatttcagcatggccaatgaggtcaaggctagattg 650
3865 TTGCTGGAGAATCCCGACTATCGCTACAATCCGCGTCGGTCTCAAGACAT 3914
651 ttgctggacaatcccgactatcgctacaatgcccgtcggtctcaggacat 700

FIGURE 6.2.1d) DNA sequence comparison of the mt**a**-1 genes from *N. crassa* (shown as the top line in the figure) and *S. fimicola.* The introns are shown in italics and underlined.

Figure 6.2.1e) shows the alignment of the *N. crassa* and *S. fimicola* mt**a**-1 amino acid sequences.

```
          .         .         .         .         .
  1 MDGNSTHPA...PNLKTTMAWSRISNQLGHWNDRKVIAIPLSDFLNTHPD 47
    |: |  |||     |: ||||| ||||||||||||:||||||||   |||
  1 MENNLMHPARTSAELRVTMAWSGISNQLGHWNDRKIIAIPLSDFTIAHPD 50
          .         .         .         .         .
 48 IQSGIIAEFKKATGEEGMFARDPESLGIMLLGPVKLFKPDSVVVDGNLFW 97
    | .||:||:|||||||||||||| |.| |||||| |||| |||||: ||||
 51 IHAGIVAEYKKATGEEGMFARDTEALEIMLLGPAKLFKADSVVVESNLFW 100
          .         .         .         .         .
 98 DPKGIHASAPKEQQ.KKAKIPRPPNAYILYRKDHHREIREQNPGLHNNEI 146
    |||||||  ||:|| ||||||||||||||||||:||||||||||||||
101 DPKGIHAETPKQQQKKKAKIPRPPNAYILYRKDHHRQIREQNPGLHNNEI 150
          .         .         .         .         .
147 SVIVGNMWRDEQPHIREKYFNMSNEIKTRLLLENPDYRYNPRRSQDIRRR 196
    |||||||||||||||:|||.|.||:| ||||:||||||| |||||||||
151 SVIVGNMWRDEQPHIRDKYFSMANEVKARLLLDNPDYRYNARRSQDIRRR 200
          .         .         .         .         .
197 VSPYLKIKLLNYDVNGNLLWGTVNAEDAALIRTHFHGVVRVEEMDDGCRI 246
    |||||||||||||||||||||||||||||||||||||||||||| |:||||
201 VSPYLKIKLLNYDVNGNLLWGTVNAEDAALIRTHFHGVVRVEETDEGCRI 250
          .         .         .         .         .
247 VCRPVAGSRKLRAAVVDTWMPRYTVDTTPVTEDDDA...QAFNFNDPLGG 293
    |||||||||||||| ||||||||||  ||.|:|:|     ||||| | |
251 VCRPVAGSRKLRAANVDTWMPRYTVDANPVSEEDEAAQGTLFNFNDTLDG 300
          .         .         .         .         .
294 AYFPLNEHLWITVNQNPPFNAPPPNPNPHLDFVHPDGMEAVVHNVQNMIA 343
     :||:||||| |  |||  .    | .  ||| ||. .:|:  |:||||
301 .FFPMNEHLWGTATQNP..SPALANIHALLDFGHPNSVQAITQNIQNMIT 347
          .         .         .
344 QVQEANEAAALTLPPPPPLRLLSLRLWLMIPLTQLSFPL 382
    ||        :|||  ||||| |:  || || || ||||
348 QVH......IMTLPLLPPLRLPSIWSWLTIPSTQRSFPL 380
```

FIGURE 6.2.1e) Amino acid sequence alignment of the mt**a**-1 proteins from *N. crassa* (shown as the top line in the figure) and *S. fimicola*. The putative DNA binding domain is shown underlined. Identical residues are denoted by vertival lines. Conservative changes are shown by either one dot or two, depending on the degree of similarity. Blank spaces indicate non-conservative amino acid differences between the two species.

Figure 6.2.1d) demonstrates that the positions of the mt**a**-1 gene introns are conserved between *N. crassa* and *S. fimicola*.. In the first intron the 5' and 3' splice sequences are conserved between the two species. In the second intron the 3' splice sequence, TAG, is conserved between the two species. The 5' splice sequence varies by one base between *N. crassa* and *S. fimicola*, GTAAGT vs GTAGGT respectively. Gaps to align the two species are in multiples of three and hence do not interrupt the open reading frame. The alignment was produced using GCG9 Bestfit.

Figure 6.2.1e) shows that the putative DNA binding domain, proposed to be an HMG box (Staben and Yanofsky, 1990), is perfectly conserved between the two species. Most of the the differences between the *N. crassa* and *S. fimicola* mt**a**-1 amino acid sequences reside in the carboxyl-terminal portion of ORFs. Philley and Staben (1994) deleted amino acids 216-220 in *N. crassa* mt**a**-1 and found that this mutation eliminated vegetative incompatibility function. Amino acids 216-220 in *N. crassa* mt**a**-1 correspond to amino acids 220-224 in *S. fimicola* mt**a**-1 and are conserved between the two species. Changing the Arginine at position 258 in *N. crassa* mt**a**-1 to Serine abolishes vegetative incompatibility function (Griffiths and Delange, 1978). There is a corresponding Arginine at position 262 in *S. fimicola* mt**a**-1. The residues mentioned seem to be important for vegetative incompatibility functions and are conserved between the two species. As *S. fimicola* contains both a mtA and a mt**a** idiomorph it seems likely that mutations leading to the loss of vegetative incompatibility function are not found at mt**a** in *S. fimicola*. The alignment was produced again using GCG9 Bestfit.

Codon usage in the *S. fimicola* mt**a**-1 ORF is typical for *Neurospora* genes.

6.2.2 Comparing Regions in *Neurospora* and *Sordaria* Outwith the Mating Type Gene ORFs.

6.2.2.1 Comparing the *N. crassa* and *S. fimicola* mtA-2/mtA-3 gene Upstream Sequences.

Ferreira *et al* (1996) identified mtA-2 and mtA-3, two divergently transcribed genes, in *N. crassa*. MtA-2 and mtA-3 may have overlapping promoters. Figure 6.2.2.1 shows the 272 bp immediately upstream from the mtA-2 ATG ORF start site compared between *N.*

crassa and *S. fimicola*. These 272 bp are 467bp upstream of the mtA-3 ATG ORF start site.

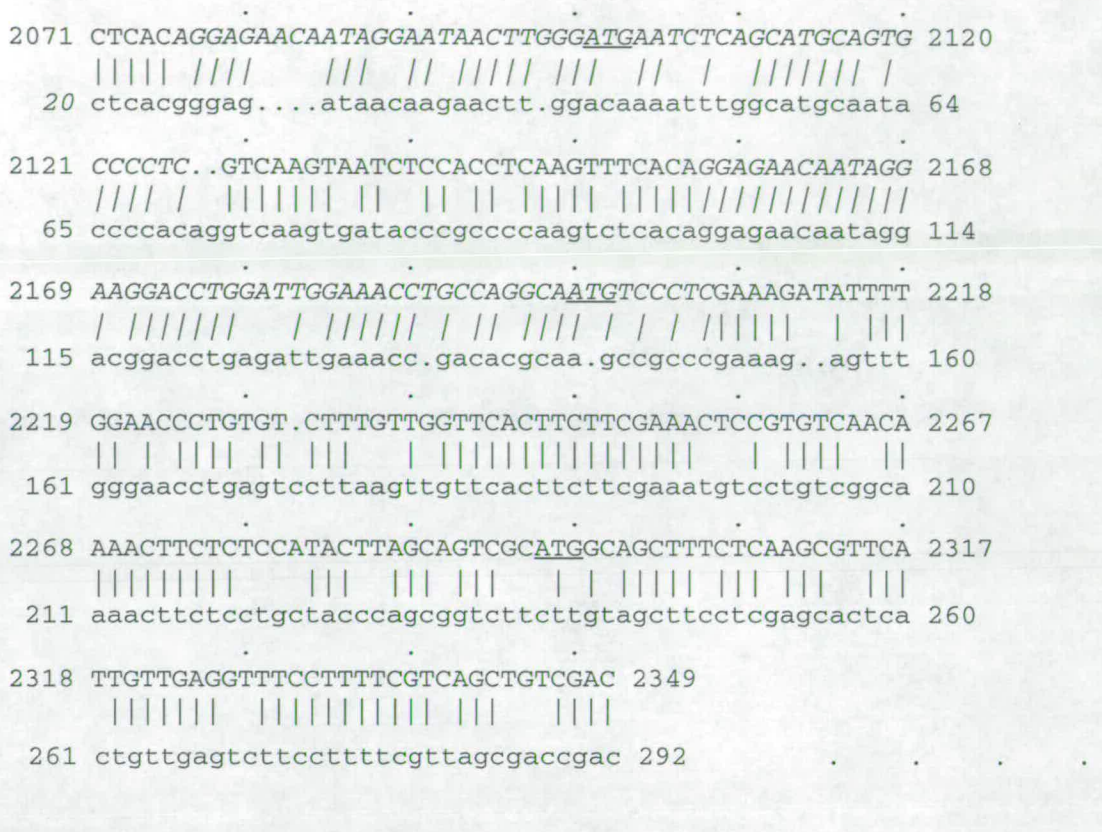


FIGURE 6.2.2.1 DNA sequence comparison of the 272 bp immediately upstream of the mtA-2 ATG ORF start site in *N. crassa* (shown as the top line) and *S. fimicola*. ATG start codons in *N. crassa* upstream of the translational start site are shown underlined. The repeated sequences in *N. crassa* are shown in italics.

Figure 6.2.2.1 shows the start codons for three untranscribed ORFs (uORFs), 22, 29 and 9 amino acids in length, present in the 5' leader sequence before the proposed translational start site for *N. crassa* mtA-2. These uORFs have been proposed by Ferreira *et al* (1996) to affect the translation of mtA-2 and mtA-3. This could provide a developmentally regulated translation system where mtA-2 and mtA-3 are transcribed constitutively but translation is restricted to post-fertilization stages. These uORFs however do not appear to be well conserved in *S. fimicola*..

The distal start sites for mtA-2 and mtA-3, not shown in the figure, are separated by only 70bp. Within these 70bp is the first 50bp sequence that contains a repeated sequence. Figure 6.2.2.1 highlights two 51bp regions present in *N. crassa* that are 64% identical with the first 50bp and contain a 15bp region that is perfectly conserved. The

transcription of mtA-2 and mtA-3 has been proposed to be coordinately controlled by specific factors and the repeated sequences may play a part in this process. In the region of *S. fimicola* that is aligned with the first 51bp sequence of *N. crassa* in the figure one can see that the 15bp repeat, ggagaacaataggaa, is not conserved in *S. fimicola* . At the second 51bp sequence in *S. fimicola* the 15bp repeat is conserved.

6.2.2.2 Analysis of the Centromere Proximal Flanking Regions

6.2.2.2.1 Comparing the Common Regions of *N. africana, N.crassa, N. sitophila, N. intermedia , N. discreta, N.tetrasperma, S. equina* and *S. sclerogenia.*

N.africana ATTTGACATGTCGACGAGAATAAA

N.crassa ATTTGACATGTCGTTGAGA TAAA

N.sitophila ATTTGACATGTCGTTGAGA TAAA

N.intermedia ATTTGACATGTCGTCGAGA TAAA

N. discreta ATTTGACATGTCGTCGAGA TAAA

N.tetrasperma ATTTGACATGTCGTCGAGA TAAA

S.equina ATTTGACATGTCGTCGAGA TAAA

S.sclerogenia ATTTGACATGTCGTCGAGA TAAA

N.africana GCTAT GGCCCGCCGCCAACGCCA

N.crassa GAAACAGGCCCGCCGCTGACGGCA

N.sitophila GAAACAGGCCCGCCGCTGACGGCA

N.intermedia GCCAT GGCC GCCGCTGACGCCA

N.discreta GCTAT GGCC GCCGCTGACGCCA

N. tetrasperma GCCAGGGGCC GCCGCTGACGCCA

S.equina GCTGT GGTCCGCCACTGACGCCA

S.sclerogenia GCTGT GGTCCGCCACTGACGCCA

N.africana ACGCTTATGCAT

N.crassa ACGCTTATGCAT

N.sitophila ACGCTTATGCAT

N. intermedia ACGCTTATGCAT

N. discreta ACGCTTATGCAT

N. tetrasperma ACGCTTATGCAT

S.equina ACGCTTATGCAT

S. sclerogenia ATGCTTATGCAT

FIGURE 6.2.2.2.1 DNA sequence alignment of the common regions of the above species.

Figure 6.2.2.2.1 shows that the *S. equina* and *S. sclerogenia* have common regions that are very well conserved when compared to the common regions of *Neurospora* species. The restriction site for *Nsi* I, ATGCAT, is perfectly conserved in all the above species. The *Nsi* I site was used by Randall and Metzenberg (1995) to identify the start of the variable region in *Neurospora* species.

6.2.2.2.2 Comparison of the *S. equina* and *S. sclerogenia* Variable Regions with the Variable Regions of *N. crassa, N. africana, N. sitophila, N. discreta, N. intermedia* and *N. tetrasperma.*

The *S. equina* variable region DNA sequence produced in this project were used for a gcg9 fasta search to see if any sequences in the database showed homology with these variable regions. Figure 6.2.2.2.2a) shows the result of the fasta search with the *S. equina* variable region sequence available.

S. equina vs *N. tetrasperma.*

```
69.4% identity in 173 bp overlap

                                1640       1650       1660       1670      1679
equinacomple                    ACAACCTCACGCTAGCGTAAATCCGTCGCTGGGCAGTACCTGC
                                                   ||| | |  |   ||| |  | ||  ||
NTMTA1AJ      TAATGGTGCAAAACACGCTTAAAGAGATGCAGCCTCACACTAGCGCGGTCCCGT--CTAT
               1540       1550       1560       1570       1580        1590

           1680       1690       1700       1710       1720       1730
equinacomple GTA-TGTACACTAGCGTAGAAGCTGTGCCCCTCATTTATGGATCTGCTAATGTTTCTTGA
             ||| ||||| |||  |||||||||||||||| |||||| |||||||| || ||| | | |
NTMTA1AJ      GTAGTGTAC-CTACAGTAGAAGCTGTGCCCTTCATTTGTGGATCTGCCAAAGTTGCCTAA
                 1600       1610       1620       1630       1640       1650

            1740       1750       1760       1770       1780      17     1790
equinacomple CCTGCAGTTTC-TGGAGATCCGCATCAGT-CCATTGGAGAAG-CCGCATACCGT-----C
             | ||| |||||  || |||||||||| || ||| |||  |||  | | | ||||      |
NTMTA1AJ      CTTGCGGTTTCGGGGGGATCCGCATCGGTCCCAGTGGGCAAGTGCCCCTCCCGTCTCTGC
                 1660       1670       1680       1690       1700       1710

                    1800       1810       1820       1830       1840       1850
equinacomple TTCATTTCACTAGTACTTCTACACAGTCGAGCCCGTTGCAAAATCGAAGTGTACTTAACA
             | | | || | || | |||||||
NTMTA1AJ      TACCTCTCCCAAGAAATTCTACATGCAGGCATGTACCCGAAAACTCCGCAGAATACCTAT
                 1720       1730       1740       1750       1760       1770
```

S. equina vs *N. intermedia*

81.5% identity in 81 bp overlap

```
          1660       1670       1680       1690       1700       1710
equinacomple CCGTCGCTGGGCAGTACCTGCGTATGTACACTAGCGTAGAAGCTGTGCCCCTCATTTATG
                                              ||| |||||||||||||| ||||||  ||
NIMTA1AH     ACACCAGCGCGGTTCCGTCTATTAGTGCACCTACAGTAGAAGCTGTGCCCTTCATTTGTG
                  1580       1590       1600       1610       1620       1630

          1720       1730       1740       1750       1760       1770
equinacomple GATCTGCTAATGTTTCTTGACCTGCAGTTTCT-GGAGATCCGCATCAGTCCATTGGAGAA
             |||||||  ||  |||  |  |  ||  |||  ||||||  ||  ||||  |||||  ||||
NIMTA1AH     GATCTGCCAAAGTTGCCTAACTTGCGGTTTCTGGGGGATCTGCATCGGTCCCAGCGGCTA
                  1640       1650       1660       1670       1680       1690

          1780       1790       1800       1810       1820       1830
equinacomple GCCGCATACCGTCTTCATTTCACTAGTACTTCTACACAGTCGAGCCCGTTGCAAAATCGA

NIMTA1AH     CCTCTCCAAGAAATTCTACATGCAGGGCATGTACCCGAAAACTCCCGCAGAATATGTGGA
                  1700       1710       1720       1730       1740       1750
```

S. equina vs *N. discreta*

71.3% identity in 122 bp overlap

```
                  1640       1650       1660       1670       1680        1689
equinacomple      ACAACCTCACGCTAGCGTAAATCCGTCGCTGGGCAGTACCTGCG---TATGTACAC
                                              ||| |  | ||  ||| |   ||||||| |
NDMTA1AG     CCAACGCTTATGCATGCAACCTCACACTAGCGCGGTTCCGTCGCTGGGCGCTATGTAC-C
                  1340       1350       1360       1370       1380       1390

          1690       1700       1710       1720       1730       1740
equinacomple TAGCGTAGAAGCTGTGCCCCTCATTTATGGATCTGCTAATGTTTCTTGACCTGCAGT---
             |  ||||||||||||||| ||||||  ||||||||| || |||  ||  || |||  ||
NDMTA1AG     CACAGTAGAAGCTGTGCCCTTCATTTGTGGATCTGCCAAAGTTGCTCAACTTGCGGTTCC
                  1400       1410       1420       1430       1440       1450

             1750       1760       1770       1780       1790       1800
equinacomple -TTCTGGAGATCCGCATCAG-TCCATTGGAGAAGCCGCATACCGTCTTCATTTCACTAGT
             ||||||  ||  ||||||  |  |||  ||||  |
NDMTA1AG     ATTCTGGGGAATCGCATCGGCCCCAGTGGACACTGGACAGCTGCCTCCCGTCTCTGCTAT
                  1460       1470       1480       1490       1500       1510

               1810       1820       1830       1840       1850       1860
equinacomple ACTTCTACACAGTCGAGCCCGTTGCAAAATCGAAGTGTACTTAACAGTCAGAGTGACTAA

NDMTA1AG     GTGCACTAGGTACAAGGATGTGCATGCGGGCATGTACCGAGAACTCCGCAGAATGTGTGG
                  1520       1530       1540       1550       1560       1570
```

S. equina vs *N. africana*

90.7% identity in 43 bp overlap

```
                1670       1680       1690       1700       1710       1720
equinacomple TCGCTGGGCAGTACCTGCGTATGTACACTAGCGTAGAAGCTGTGCCCCTCATTTATGGAT
                                                | ||||||||||||||||||||||||||
NAMTA1AA     ATCACGCTAACCTTTCTGATCCCGGTAGAGGTGTAGAAGCTGTGCCCCTCATTTATGGAT
               610        620        630        640        650        660

                1730       1740       1750       1760       1770       1780
equinacomple CTGCTAATGTTTCTTGACCTGCAGTTTCTGGAGATCCGCATCAGTCCATTGGAGAAGCCG
             |||| || ||| |
NAMTA1AA     CTGCCAAAGTTGCCCTCCCAGTGGGCAGCTGCCTCCCGTCTCTACAAGGAATGTACATAC
               670        680        690        700        710        720
```

FIGURE 6.2.2.2.2a) The results of the fasta search using the partly sequenced *S. equina* variable region. The above figure shows the four best scores for homology with the *S. equina* variable region were sequences from the variable regions of *N. tetrasperma, N. intermedia, N. discreta* and *N. africana*

In figure 6.2.2.2.2a) one can see that a small part of the variable region of *S. equina* shows homology with a small part of the variable region in *N. tetrasperma, N. intermedia, N. discreta* and *N. africana.* Randall and Metzenberg (1995) demonstrated that the variable regions of the mtA mating types of *N. crassa* and *N. sitophila* are very similar to one another but completely dissimilar to the variable regions of the other *Neurospora* species included in the study. The variable regions of the mtA mating types of *N. intermedia* and *N. tetrasperma* are similar. *N. africana, N. intermedia* and *N. discreta* have an "island of homology" ~300-400 nucleotides long and 50-60% similar to one another. Different islands of homology are present between individual pairs of variable regions. It is interesting that the fasta search using the variable region of *S. equina* placed this Sordaria species with the *N. intermedia, N. tetrasperma, N. discreta* and *N. africana* group and no homology was found with the *N. crassa* and *N. sitophila* group. It seems that *S. equina* shares an island of homology with these species. The functional significance of the variable region and the islands of homology is not known.

The sequenced portions of the variable regions of *S. equina* and *S. sclerogenia* were compared using GCG9 bestfit. The result of the comparison is shown in figure 6.2.2.2.2 b).

S. sclerogenia vs *S.equina*

94% identity in 70bp overlap

```
1814 cctgcgtatgtacgctagtgtagaagccgtgcccctcatttatggatct. 1862
     |||||||||||||| |||| |||||||| ||||||||||||||||||||
1675 cctgcgtatgtacactagcgtagaagctgtgcccctcatttatggatctg 1724
                .          .
1863 ....tgttccttgacctgcag 1879
         |||| ||||||||||||
1725 ctaatgtttcttgacctgcag 1745
```

FIGURE 6.2.2.2.2b) DNA sequence comparison of the variable region of *S. sclerogenia* (top line) and the variable region of *S. equina.*

Figure 6.2.2.2.2b) shows that the region with which *S. equina* shows homology to the variable regions of the *Neurospora* species shown in figure 6.2.2.2.2a) is also the region with which *S. equina* shows high homology to the variable region of *S. sclerogenia. S. sclerogenia* shows no sequence homology with the variable regions of *N. crassa* and *N. sitophila.*

6.2.2.3 Analysis of the Centromere Distal Flanking Region in *S. fimicola* mta.

Some sequence data was obtained for a region downstream of the end of the *S. fimicola* mt**a**-1 gene. The centromere distal flank starts 216 bp downstream of the end of mt**a**-1 in *N. crassa.* The sequence data obtained from *S. fimicola* mt**a** was compared to the *N. crassa* mt**a** sequence using GCG9 bestfit. The result of this analysis is shown in figure 6.2.2.3.

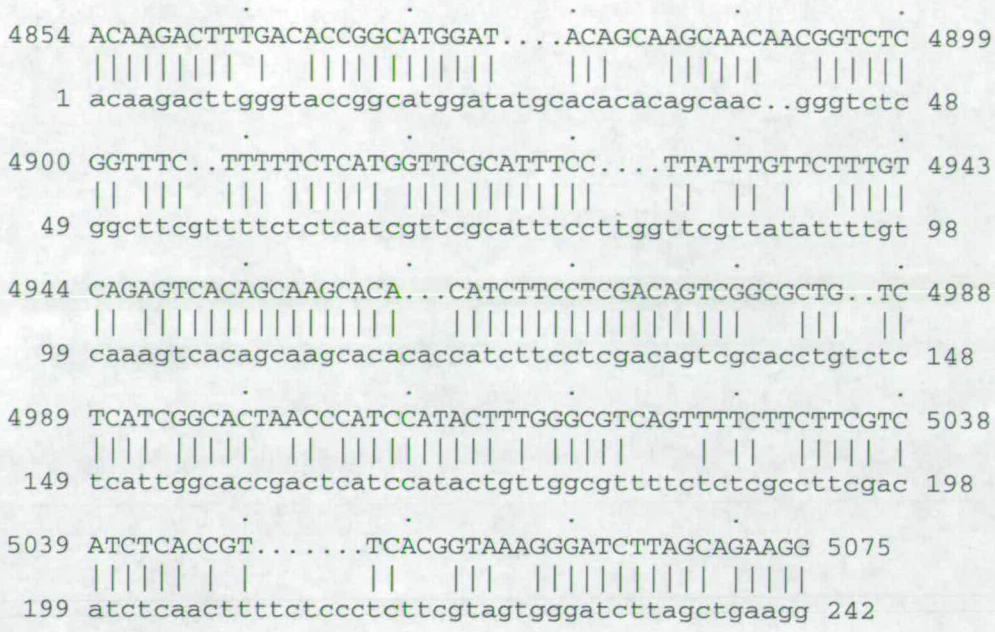


FIGURE 6.2.2.3 DNA sequence comparison of *N. crassa* mt**a** centromere distal sequences (top line) vs *S. fimicola* mt**a** sequences.

Figure 6.2.2.3 demonstrates that the *S. fimicola* sequence downstream of the mt**a**-1 gene shows high homology (81%) to centromere distal flanking sequence in *N. crassa* mt**a**. The mt**a** idiomorph in *S. fimicola* and the mt**a** idiomorph in *N. crassa* appear to be located at equivalent centromere distal chromosomal locations.

6.3 Discussion

The phylogeny shown in figure 6.2.1c) supports the tree constructed by Randall and Metzenberg (1995) in that *N. crassa* and *N. sitophila* are more closely related to each other than to other species and *N. intermedia* and *N. tetrasperma* are related. This supports the evidence from the variable region data which also suggested a close relationship between *N. crassa* and *N. sitophila*. Figure 6.2.1c) also supports the proposal that the pseudohomothallic *N. tetrasperma* arose after the divergance of the *N. crassa, N.intermedia, N. sitophila* and *N.tetrasperma* subgroup from its common ancestor with the heterothallic *N. discreta* (Randall and Metzenberg, 1995).

There have been previous studies to try and resolve the evolutionary history of the *Neurospora* genus using data from other genes. Taylor and Natvig (1989) examined the distribution of restriction-endonuclease sites among mitochondrial DNAs (mtDNAs) to construct a phylogeny. Natvig *et al* (1987) used restriction-site analysis of anonymous nuclear-DNA fragments to infer evolutionary relationships among *Neurospora* species. Most recently Skupski *et al* (1997) used restriction fragment polymorphisms derived from cosmid probes and sequence data from the upstream regions of two genes, al-1 and frq to build a phylogeny for five heterothallic *Neurospora* species. Maximum likelihood and parsimony trees were constructed. All the studies, including that of Randall and Metzenberg (1995) place *N. discreta* relatively distant from the other *Neurospora* species. Trees based on mtDNAs and the Skupski *et al* (1997) study place *N. crassa* and *N. intermedia* as sister taxa and may even be incompletely resolved sister taxa. All the trees in previous studies except for that of Randall and Metzenberg (1995) and the frq tree support *N. tetrasperma* and *N. sitophila* as the closest relatives of one another. Using a Maximum Likelihood test to compare the frq and mtA-1 trees, Skupski *et al* (1997) showed that these two trees are significantly different. There is evidence that different genes have different evolutionary histories leading to conflicting trees.

One of the cosmids used in the Skupski *et al* (1997) study, 16:4F, mapped near the centromere on the right arm of linkage group I. The mating type locus maps to the left arm of this linkage group. Using the restriction fragment data from 16:4F, trees were built to establish whether they would give a phylogeny similar to that based on the mtA-1 data. This was not the case, only the mtA-1 tree groups *N. crassa* with *N. sitophila* and *N. tetrasperma* with *N. intermedia*.. Similarity between these species is limited apparently to the mating type idiomorph.

How can different genes have different evolutionary histories? The most plausible theory is that of ancestral polymorphism where one must realize that a node in a tree represents a group of individuals among which variation can exist. The polymorphism in a group of individuals could have existed before the evolution of the population from a distant ancestor. This of course can result in inaccurate phylogenetic trees. Again it must be stressed that inaccuracies can also occur as a result of limited data being available from which to construct a phylogeny.

The shift from a well conserved common region to a dissimilar variable region is intriguing. No function has been assigned to the common and variable regions. Perhaps the common region may have a role as a pairing site for homologous chromosomes

during meiosis (Randall and Metzenberg, 1995). The variable region may have arisen by a similar mechanism to which the mtA and mt**a** idiomorphs evolved, they too are regions of complete dissimilarity. The variable region could be important in the process of speciation and as homothallic variable regions appear to be conserved from homothallic species to species, the variable region may be part of the distinction between homothallism and heterothallism. Randall and Metzenberg (1995) suggested that a study to try and establish the existance of other variable regions at other loci within the *Neurospora* genus may identify species-specific genes in these variable regions.

In conclusion the mating type genes of *Neurospora* and *Sordaria* species are conserved reflecting their important role in the sexual cycle as transcription factors. A homothallic *S. equina* mtA gene is functional with respect to mating type activity in a mutant *Neurospora* strain but cannot confer homothallic behaviour on the recipient indicating that it is not a change to the mtA gene itself that results in homothallic behaviour. The evolutionary relationships between homothallic and heterothallic species have yet to be conclusively established.

Chapter 7

Further Work

7 Further Work

Ideally with more time more *Sordaria* species could have been sequenced to be included in the analysis. Of particular interest would have been the sequencing of the mtA idiomorph of *S. brevicollis* which is a heterothallic species. MtA individuals in this species show homothallic behaviour at a low frequency (Robertson *et al*, manuscript submitted to Mycological Research). Sequencing and expression of the mtA-1 gene would show if it is changes to this gene that are responsible for the homothallic behaviour displayed.

The inclusion of more species would have led to a more informative phylogeny for establishing evolutionary relationships. Sequence data from *N. terricola* would be of interest to see what the relationship is between the mtA and mt**a** idiomorphs in that species. A λ clone is available for *S. macrospora* which hybridises to both mating type probes indicating that in this species the mtA and mt**a** mating type idiomorphs are linked. Sequencing the junction between the idiomorphs in *S. macrospora* would show if they have been recombined in a similar way to the idiomorphs in *S. fimicola*. Obviously comparing *Sordaria* species to *Neurospora* species to find a region of homology that could result in unequal crossing over is not ideal. Obtaining more sequencing data from *Sordaria* species could reveal a site of homology between the two idiomorphs which might have led to unequal crossing over.

The ability of *S. equina* mtA to confer vegetative incompatibility in *N. crassa* was not tested due to the unexplained loss of the *ad-3B*. forcing marker. It would be of interest to establish whether *S. equina* does have vegetative incompatibility function bearing in mind that this species never comes into contact with a mt**a** idiomorph. Testing the *S. fimicola* mtA and mt**a** idiomorphs for vegetative incompatibility function by transformation into a sterile heterothallic species would be worthwhile considering that both idiomorphs reside in the one nucleus.

References

Aber W., Enquist L., Hohn B., Murray N. E., Murray K. (1983) "Experimental methods for use with Lambda " **Lambda II p433-466, Cold Spring Harbor.**

Arnaise S., Zickler D., Glass N. L. (1993) " Heterologous expression of mating-type genes in filamentous fungi " **Proc.Natl. Acad. Sci. USA 90: 6616-6620.**

Astell C.R., Ahlstrom-Jonasson L., Smith M. (1981) " The Sequence of the DNAs Coding for the Mating-Type Loci of *Saccharomyces cerevisiae* " **Cell 27: 15-23.**

Aurora R., Herr W. (1992) " Segments of the POU domain influence one anothers DNA binding specificity" **Mol. Cell. Biol. 12: 455-467.**

Banham A.H., Asante-Owusu R.N., Gottgens B., Thompson S.A.J., Kingsnorth C.S., Mellor E.J.C., Casselton L.A. (1995) " An N-Terminal Dimeriztion Domain Permits Homeodomain Proteins to Choose Compatible Partners and Initiate Sexual development in the Mushroom *Coprinus cinereus* " **The Plant Cell 7: 773-783.**

Banuett F. (1992) "*Ustilago maydis,* the delightful blight " **TIG 8: 174-180.**

Beadle G. W., Coonradt V. L. (1944) " Heterokaryosis in *Neurospora crassa* " **Genetics 29: 291-308.**

Beatty N. P., Smith M. L., Glass N. L. (1994) " Molecular characterisation of mating-type loci in selected homothallic species of *Neurospora, Gelasinospora* and *Anixiella* " **Mycol. Res. 98 (11): 1309-1316.**

Bender A., Sprague G.F. (1987) " MATα1 Protein, a Yeast Transcriptional Activator, Binds Synergistically with a Second Protein to a Set of Cell-Type Specific Genes" **Cell 50: 681-691.**

Benton W., Davis R. (1977) " Screening λgt recombinant clones by hybridisation to single plaques *in situ* " **Science 196: 180-182.**

Betz R., Crabb J.W., Meyer H.E., Wittig R., Duntze W. (1987) " Amino acid sequences of a-factor mating peptides from *Saccharomyces cerevisiae* " **J.Biol.Chem. 262: 546-548.**

Birnboim H.C., Dolly J. (1979) " A rapid alkaline extraction procedure for screening recombinant plasmid DNA" **Nuc.Acids Res. 7: 1513.**

Bistis G. N. (1981) " Chemotropic Interactions Between Trichogynes and Conidia of Opposite Mating-Type in *Neurospora crassa* " **Mycologia 73: 959-975.**

Bistis G. N. (1983) " Evidence for diffusible, mating-type specific trichogyne attractants in *Neurospora crassa* " **Exp.Mycol. 7: 292-295**

Bolker M., Urbin M., Kahmann R. (1992) " The a mating type of *Ustilago maydis* specifies cell signalling components" **Cell 68: 441-450.**

Borges M. I., Azevedo M. O., Bonatelli R., Felipe M.S.S., Astolfi-Filho S. (1990) "A practical method for the preparation of total DNA from filamentous fungi" **Fungal Genetics News Letter 37: 10.**

Bruhn L. Sprague G.F. (1994) " MCM1 point mutations deficient in expression of α-specific genes: residues important for interaction with α1" **Mol. Cell. Biol. 14: 253-**

Buller A.H.R (1931) "Researches on fungi, vol IV" **Longmans Green, London.**

Bullock W.O., Fernandez T.M., Short J.M. (1987) " XL-1 Blue-A high efficiency plasmid transforming Reca *Escherichia coli* strain with Beta-galactosidase selection" **Biotechniques 5: 376-378.**

Casselton L.A. (1978) " Dikaryon formation in higher basidiomycetes" **The Filamentous Fungi Vol. 3, Chapter 14 pp. 275-297.**

Casselton L.A., Asante-Owusu R.N., Banham A.H., Kingsnorth C.S., Kues U., O'Shea S.F., Pardo E.H. (1995). " Mating type control of sexual development in *Coprinus cinereus* " **Can. J. Bot. 73: S266-S272.**

Casselton L.A., Kues U. (1994) "Mating type genes in homobasidiomycetes" **The Mycota, Sringer-Verlag, Heidelberg. Chap.18, pp 307-321.**

Chang S., Staben C. (1994) " Directed Replacement of mtA by mta-1 Effects a mating Type Switch in *Neurospora crassa* " **Genetics 138: 75-81**

Day P.R. (1960) " The structure of the A mating type locus in *Coprinus lagopus* " **Genetics 45: 641-650.**

Day P.R. (1963) " Mutations affecting the A mating-type locus in *Coprinus lagopus* " **Genet. Res. Camb. 4: 55-65.**

Day P.R., Anagnostakis S.L., Puhalla J.E. (1971) " Pathogenicity resulting from mutation at the b locus of *Ustilago maydis* " **Proc. Natl. Acad. Sci. USA 68: 533-535.**

Debuchy R., Arnaise S., Lecellier G. (1993) " The mat- allele of *Podospora anserina* contains three regulatory genes required for the development of fertilised female organs " **Mol. Gen. Genet. 241: 667-673.**

Debuchy R., Coppin E. (1992) "The mating types of *Podospora anserina :* functional analysis and sequence of the fertilisation domains " **Mol. Gen. Genet. 233: 113-121.**

Dolan J.W., Fields S. (1991) "Cell-type-specific transcription in yeast" **Biochim. Biophys. Acta. 1088: 155-169.**

Elble R., B-K. Tye (1991) " Both activation and repression of a-mating-type-specific genes in yeast require transcription factor MCM1" **Proc. Natl. Acad. Sci. USA 88: 10966-10970.**

Errede B., Ammerer G. (1989) "STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes" **Genes Dev. 3: 1349-1361.**

Felsenstein J., Churchill G.A. (1996) " A hidden markov model approach to variation among sites in the rate of evolution." **Molecular Biology and Evolution 13: 93-104.**

Ferreira A.V.B, Saupe S., Glass N.L. (1996) "Transcriptional analysis of the mtA idiomorph of *Neurospora crassa* identifies two genes in addition to mtA-1" **Mol. Gen. Genet. 250: 767-774.**

Fields S., Herskowitz I. (1987) "Regulation by the Mating-Type Locus of STE12, a Gene Required for Cell-Type-Specific Expression" **Mol.Cell. Biol. 7: 3818-3821.**

Fincham J.R.S., Day P.R., Radford A. (1979) " Fungal Genetics " **Blackwell Scientific Publications p24-32 .**

Fincham J.R.S (1983) "Genetics" **Jones and Bartlett Publishers Inc. p55.**

Frischauf A-M., Lehrach H., Poustka A., Murray N. (1983) " Lambda Replacement Vectors Carrying Polylinker Sequences " **J. Mol. Biol. 170: 827.**

Garnjobst L., Wilson J. F. (1956) " Heterokaryosis and Protoplasmic Incompatibility in *Neurospora crassa* " **Proc. Natl. Acad. Sci. USA 42: 613-618.**

Gillissen B., Bergemann J., Sandmann C., Schroeer B., Bolker M., Kahmann R. (1992) " A Two Component Regulatory System for Self/Non-Self Recognition in *Ustilago maydis* " **Cell 68: 647-657.**

Glass N. L., Grotelueschen J., Metzenberg R. L. (1990a) " *Neurospora crassa* A mating-type region " **Proc. Natl. Acad. Sci. USA. 87: 4912-4916.**

Glass N. L., Kuldau G. A. (1992) " Mating Type and Vegetative Incompatibilty in Filamentous Ascomycetes " **Annu. Rev. Phytopathol. 30: 201-224.**

Glass N. L., Lee L. (1992) "Isolation of *Neurospora crassa* A Mating Type Mutants by Repeat Induced Point (RIP) Mutation" **Genetics 132: 125-133.**

Glass N. L., Metzenberg R. L., Raju N. B. (1990b) " Homothallic Sordariaceae from Nature: The Absence of Strains Containing Only the a Mating Type Sequence" **Experimental Mycology 14: 274-289.**

Glass N. L., Smith M. L. (1994) " Structure and function of a mating-type gene from the homothallic species *Neurospora africana* " **Mol. Gen. Genet. 244: 401-409.**

Glass N. L., Vollmer S. J., Staben C., Grotelueschen J., Metzenberg R. L., Yanofsky C. (1988) "DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar " **Science 241: 570-573.**

Goutte C., Johnson A.D. (1993) " Yeast a1 and α2 homeodomain proteins form a DNA-binding activity with properties distinct from either protein" **J. Mol. Biol. 233: 359-371.**

Griffiths A.J.F. (1982) " Null mutants of the A and a mating-type alleles of *Neurospora crassa* " **Canadian Journal of Genetics and Cytology 24: 167-176.**

Griffiths A.J.F., Delange A.M. (1978) " Mutations of the a mating type in *Neurospora crassa* " **Genetics 88: 239-254.**

Hagen D.C., Sprague G.F.(1984) " Induction of the yeast α-specific STE3 gene by the peptide pheromone a-factor " **J.Mol. Biol. 178: 835-852.**

Hartig A., Holly J., Saari G., MacKay V.L. (1986) "Multiple Regulation of STE2, a Mating-Type-Specific Gene of *Saccharomyces cerevisiae* " **Mol. Cell. Biol. 6: 2106-2114.**

Herskowitz I (1988) "The Life Cycle of the Budding Yeast *Saccharomyces cerevisiae* " **Microbiological Reviews 52: 536-553.**

Hicks J.B., Strathern J.N., Klar A.J.S. (1979) "Transposable mating type genes in *Saccharomyces cerevisiae* " **Nature 282: 478-483.**

Holliday R. (1961) " The genetics of *Ustilago maydis* " **Genet. Res. Camb. 2: 204-230.**

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

Huelsenbeck J. P. (1995) "Performance of phylogenetic methods in simulation" **Syst. Biol. 44: 17-48.**

Ingraham H.A., Flynn S. E., Voss J. W., Albert M. S., Kapiloff M. S., Wilson L., Rosenfeld M. G. (1990) "The POU-specific domain of PIT-1 is essential for sequence specificity, high affinity DNA binding and DNA dependent PIT-1-PIT-1 interactions" **Cell 61: 1021-1033.**

Jantzen H-M., Admon A., Bell S. P., Tjion R. (1990) " Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins " **Nature 234: 113-121.**

Jarvis E.E., Hagen D.C., Sprague G.F. (1987) " Identification of a DNA segment that is necessary and sufficient for α-specific gene control in *Saccharomyces cerevisiae* : implications for regulation of α-specific and a-specific genes" **Mol. Cell. Biol. 8: 309-320.**

Jenness D.D., Burkholder A.C., Hartwell L.H. (1983) " Binding of α-factor pheromone to yeast a-cells: chemical and genetic evidence for an α–factor receptor " **Cell 35: 521-529.**

Johnson A.D. (1995) " Molecular mechanisms of cell-type determination in budding yeast" **Current Opinion in Genetics and Development 5: 552-558.**

Kamper J., Reichmann M., Romeis T., Bolker M., Kahmann R. (1995) " Multiallelic Recognition: Nonself-Dependent Dimerization of the bE and bW Homeodomain Proteins in *Ustilago maydis* " **Cell 81:73-83.**

Keleher C.A., Goutte C., Johnson A.D. (1988) " The yeast cell-type-specific repressor α2 acts cooperatively with a non-cell-type-specific protein" **Cell 53: 927-936.**

Keleher C.A., Redd M.J., Schultz J., Carlson M., Johnson A.D. (1992) " SSN6/TUP1 is a general repressor of transcription in yeast" **Cell 68: 709-719.**

Kissinger C.R., Liu B., Martin-Blanco E., Kornberg T.B., Pabo C. O. (1990) " Crystal structure of an engrailed homeodomain-DNA complex at 2.8A resolution- a framework for understanding homeodomain-DNA interactions" **Cell 63: 579-590.**

Kronstad J.W. , Leong S. A. (1989) " Isolation of two alleles of the b locus of *Ustilago maydis* " **Proc. Natl. Acad. Sci. USA 86:978-982.**

Kronstad J.W., Leong S.A. (1990) " The b mating type locus of *Ustilago maydis* contains variable and constant regions" **Genes and Development 4: 1384-1395.**

Kues U., Asante-Owusu R.N., Mutasa E.S., Tymon A.M., Pardo E.H., O'Shea S.F., Gottgens B., Casselton L.A. (1994c) " Two Classes of Homeodomain Proteins Specify the Multiple A Mating Types of the Mushroom *Coprinus cinereus* " **Plant Cell 6: 1467-1475.**

Kues U., Casselton L.A. (1992) " Fungal mating type genes-regulators of sexual development " **Mycol. Res. 96 (12): 993-1006.**

Kues U., Casselton L.A. (1993) " The origin of multiple mating types in mushrooms" **Journal of Cell Science 104: 227-230.**

Kues U., Gottgens B., Stratmann R., Richardson W.V.J., O'Shea S.F., Casselton L.A. (1994a). " A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus* ". **EMBO J. 13:4054-4059.**

Kues U., Richardson W.V.J., Tymon A.M., Mutasa E.S., Gottgens B., Gaubatz S., Gregoriades A., Casselton L.A. (1992) " The combination of dissimilar alleles of the Aα and Bβ gene complexes, whose proteins contain homeo domain motifs, determines sexual development in the mushroom *Coprinus cinereus* " **Genes and Development 6: 568-477.**

Kues U., Tymon A.M., Richardson W.V.J., May G., Gieser P.T., Casselton L.A. (1994b) " A mating-type factors of *Coprinus cinereus* have variable numbers of specificity genes encoding two classes of homeodomain proteins" **Mol. Gen. Genet. 245: 45-52.**

Liddle S. (1993) " Analysis of the mating-type locus of two species of *Sordaria:* An investigation of homothaliic development/evolution " **Unpublished material.**

MacKay V.L., Manney T.R. (1974) "Mutations affecting sexual conjugation and related processes in *Saccharomyces cerevisiae* . II Genetic analysis of nonmating mutants" **Genetics 76: 273-288.**

Mak A., Johnson A.D. (1993) " The carboxy-terminal tail of the homeodomain protein α2 is required for function with a second homeodomain protein" **Genes Dev. 7: 1862-1870.**

Maniatis J., Fritsch E.F., Sambrook J. (1989) " Molecular cloning: a laboratory manual " **Cold Spring Harbour , New York: Cold Spring Harbour Laboratories.**

May G., Le Chevanton L., Pukkila P.J. (1991) " Molecular analysis of the *Coprinus cinereus* Mating Type A Factor Demonstrates an Unexpectedly Complex Structure" **Genetics 128:529-538.**

Metzenberg R. L., Glass N. L.(1990)"Mating Type and Mating Strategies in *Neurospora*" **Bioessays 12: 53-59.**

Mutasa E.S., Tymon A.M., Gottgens B., Mellon F.M., Little P.F.R., Casselton L.A. (1990) " Molecular organisation of an A mating type factor of the basidiomycete fungus *Coprinus cinereus* " **Current Genetics 18; 223-229.**

Nasmyth K.A., Tatchell K. (1980) " The structure of transposable yeast mating type loci" **Cell 19: 753-764.**

Natvig D.O., Jackson D. A., Taylor J. W. (1987) " Random fragment hybridisation analysis of evolution in the genus *Neurospora* : the status of four-spored strains" **Evolution 41: 1003-1021.**

Nelson M. A., Metzenberg R. L. (1992) "Sexual Development Genes of *Neurospora crassa* " **Genetics 132: 149-162.**

Orbach M.J., Sachs M.J., Yanofsky C. (1990) "*N. crassa arg-2* locus" **J. Biol. Chem. 265: 10981-10987.**

Pall M. L., Brunelli J. P. (1993) " A series of six compact fungal transformation vectors containing polylinkers with multiple unique restriction sites " **Fungal Genetics Newsletter 40: 59-62.**

Pall M. L., Brunelli J. P. (1994) " New plasmid and λ / plasmid hybrid vectors and a Neurospora crassa genomic library containing the *bar* selectable marker and the *Cre//lox* site-specific recombination system for use in filamentous fungi " **Fungal Genetics Newsletter 41: 63-65.**

Paluh J.L., Orbach M.J., Legerton T.L., Yanofsky C. (1988) "The cross pathway control gene of *N. crassa, cpc-1,* encodes a protein similar to GCN4 of yeast and DNA binding domain of oncogene v-jun encoded protein" **Proc. Natl. Acad. Sci. USA 85: 3728-3732.**

Philley M., Staben C. (1994) " Functional Analysis of the *Neurospora crassa* MT a-1 Mating Type Polypeptide " **Genetics 137: 715-722.**

Picard M., Debuchy R., Coppin E. (1991) "Cloning the Mating Types of the Heterothallic Fungus *Podospora anserina:* Developmental Features of Haploid Transformants Carrying Both Mating Types " **Genetics 128: 539-547.**

Primig H., Winkler H., Ammerer G. (1991) " The DNA binding and oligomerization domain of MCM1 is sufficient for its interaction with other regulatory proteins" **EMBO J. 10: 4209-4218.**

Puhalla J.E. (1970) " Genetic studies of the b incompatibility locus of *Ustilago maydis.*" **Genet. Res. Camb. 16: 229-232.**

Raju N. B. (1977) " Meiotic nuclear behaviour and ascospore formation in five homothallic species of *Neurospora* " **Canadian Journal of Botany 56: 754-763.**

Randall T.A., Metzenberg R.L. (1995) " Species-Specific and mating Type-Specific DNA Regions Adjacent to Mating Type Idiomorphs in the Genus *Neurospora*" **Genetics 141: 119-136.**

Raper J.R. (1966) " Genetics of Sexuality in Higher Fungi" **The Ronald Press, New York.**

Richardson W.V.J., Kues U., Casselton L.A. (1993). "The A mating-type genes of the mushroom *Coprinus cinereus* are not differentially transcribed in monokaryons and dikaryons" **Mol. Gen. Genet. 238: 304-307.**

Rine J., (1986) "Regulation of cell division by peptide hormones of *S. cerevisiae* " **TIG 2: 276-277.**

Rowell J.B., DeVay J.E. (1954) " Genetics of *Ustilago zeae* in relation to basic problems of its pathogenicity" **Phytopathology 44: 356-362.**

Sanger F., Nicklen S., Coulson A. (1977) "DNA sequencing with chain-terminating inhibitors" **Proc. Natl. Acad. Sci. USA 74: 5463-5467.**

Saupe S., Stenberg L., Shiu K.T., Griffiths A.J.F., Glass N.L. (1996) "The molecular nature of mutations in the mtA-1 gene of *Neurospora crassa* A idiomorph and their relation to mating-type function" **Mol. Gen. Genet. 250: 115-122.**

Schulz B., Banuett F., Dahl M., Schlesinger R., Schafer W., Martin T., Herskowitz I., Kahmann R. (1990) " The b Alleles of *U. maydis* , Whose Combinations Program Pathogenic Development, Code for Polypeptides Containing a Homeodomain-Related Motif " **Cell 60: 295-306.**

Scott M.P., Tamkum J.W., Hartzell G.W. III (1989) " The structure and function of the homeodomain " **Biochim.Biophys.Acta 989: 25-48.**

Selker E. U. (1990) "Premeiotic instability of repeated sequences in *Neurospora crassa* " **Ann. Review Genet. 24: 579-613.**

Shear C.L., Dodge B.O. (1927) " Life histories and heterothallism of the red bread-mold fungi of the *Monilia sitophila* group. " **J.Agric.Res. 34: 1019-1042.**

Skupski M.P., Jackson D.A., Natvig D.O (1997) " Phylogenetic analysis of heterothallic *Neurospora* species" **Fungal Genetics and Biology 21:153-162.**

Southern E. M. (1975) "Detection of specific sequences among DNA fragments separated by gel electrophoresis " **J. Mol. Biol. 98: 503.**

Sprague G.F., Blair L.C., Thorner J. (1983a) " Cell Interactions and Regulation of Cell type in the Yeast *Saccharomyces cerevisiae* " **Ann. Rev. Microbiol. 37 : 623-660.**

Sprague G.F., Jensen R., Herskowitz I. (1983b) " Control of Yeast Cell Type by the Mating Type Locus: Positive Regulation of the α–specific STE3 Gene by the MATα1 Product" **Cell 32: 409-415.**

Staben C., Yanofsky C. (1990) " *Neurospora crassa* a mating-type region " **Proc. Natl. Acad. Sci. USA. 87: 4917-4921.**

Stotzler D., Kiltz H.H., Duntze W. (1976) " Primary structure of α factor peptides from *Saccharomyces cerevisiae* " **Eur.J. Biochem. 19: 397-400.**

Strathern J., Hicks J., Herskowitz I. (1981) " Control of Cell Type in Yeast by the Mating Type Locus, the α1-α2 Hypothesis" **J.Mol.Biol. 147: 357-372.**

Swiezynski K.M., Day P.R. (1960) " Heterokaryon formation in *Coprinus lagopus* " **Gen. Res., Camb. 1:114-128.**

Swofford D., Olsen G., Waddell P., Hillis D. (1996) " Phylogenetic inference. In *Molecular Systematics* " ed. Hillis D., Morritz C. **Sinauer pp. 407-514.**

Tatchell K., Nasmyth K.A., Hall B.D. (1981) " *In Vitro* Mutation analysis of the Mating-Type Locus in Yeast" **Cell 27: 25-35.**

Taylor J. W., Natvig D. O. (1989) "Mitochondrial DNA and evolution of heterothallic and pseudohomothallic *Neurospora* species" **Mycol. Res. 93: 257-272.**

Thompson J., Higgins D., Gibson T. (1994) "CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice" **Nuc. Acids Res. 22: 4673-4680.**

Thompson-Coffe C., Zickler D. (1994) " How the Cytoskeleton Recognises and Sorts Nuclei of Opposite Mating Type during the Sexual Cycle in Filamentous Ascomycetes " **Developmental Biology 165: 257-271.**

Tymon A.M., Kues U., Richardson W.V.J., Casselton L.A. (1992) "A mushroom mating-type protein that regulates sexual and aesexual development contains a POU-related domain" **EMBO J. 11: 1805-1816.**

Vogel H. J (1956) "A convenient growth medium for *Neurospora crassa* (medium N)" **Microb.Genet.Bull. 13: 42-43.**

Vollmer S.J., Yanofsky C. (1986) "Efficient cloning of genes in *Neurospora crassa* " **Proc.Natl.Acad.Sci.USA 83: 4869-4873.**

Westergaard M., Mitchell H. K. (1947). " *Neurospora* V. A synthetic growth medium favouring sexual reproduction" **Am. J. Bot. 34: 573-577.**

Wolberger C., Vershon A.K., Liu B., Johnson A.D, Pabo C.O. (1991) " Crystal-structure of a matα-2 homeodomain-operator complex suggests a general mode for homeodomain-DNA interactions" **Cell 67: 517-528.**

Xu S., Yoell H. J., Anderson J. B. (1994) "An efficient protocol for isolating DNA from higher fungi " **Technical Tips TIG 10 (7): 226.**

Yanisch-Perron C., Viera J., Messing J. (1985) " Improved M13 phage cloning vectors and host strains- nucleotide sequences of the M13MP18 and PUC19 vectors" **Gene 33: 103-119.**

Yee A.R., Kronstrad J.W.(1993) "Construction of chimeric alleles with altered specificity at the b incompatibility locus of *Ustilago maydis* "**Proc.Natl.Acad.Sci.USA 90; 664-668.**

Yuan Y.O. Stroke I.L., Fields S. (1993). " Coupling of cell identity to signal response in yeast: interaction between the α1 and STE12 proteins" **Genes Dev. 7: 1584-1597.**

Zickler D., Arnaise S., Coppin E., Debuchy R., Picard M. (1995) " Altered Mating-Type Identity in the Fungus *Podospora anserina* Leads to Selfish Nuclei, Uniparental Progeny, and Haploid Meiosis " **Genetics 140: 493-503.**