CHARACTERIZATION OF GENES PUTATIVELY INVOLVED IN MATING-TYPE ASSOCIATED VEGETATIVE INCOMPATIBILITY IN *Neurospora crassa*

by

Megan Diane Hiltz

B.Sc., Mount Allison University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
Department of Botany

and

The Biotechnology Laboratory

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 2001

In presenting this thesis in partial fulfilment of the requirements
for an advanced degree at the University of British Columbia, I
agree that the Library shall make it freely available for reference
and study. I further agree that permission for extensive copying of
this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that
copying or publication of this thesis for financial gain shall not
be allowed without my written permission.

Department of _Botany & The Biotechnology   Laboratory_

The University of British Columbia
Vancouver, Canada

Date _Sept. 17/01_

Abstract

Vegetative incompatibility is a ubiquitous phenomenon in many filamentous fungi. *Neurospora crassa* vegetative incompatibility involves the prevention of vigorous heterokaryons, mediated by genetic dissimilarity at eleven heterokaryon incompatibility (*het*) loci. During the vegetative phase, the mating-type (*mat*) locus functions as a *het* locus, while in the sexual phase, the *mat* locus confers mating identity and post-fertilization functions for completion of the sexual cycle. *tol* is an unlinked recessive mutation that suppresses mating-type vegetative incompatibility. TOL encodes a 1044 amino acid polypeptide that contains putative protein-protein interaction domains.

TOL contains a region called the SET domain that has similarity to HET-6 in *N. crassa* and HET-E in *Podospora anserina* (SET = HET-<u>S</u>IX, HET-<u>E</u> and <u>T</u>OL). Searches of the *N. crassa* genome yielded a large number of TOL-like sequences, which also share the SET domain and may indicate a novel motif region. BLASTP searches with TOL have identified similarity to Mms21p, a DNA repair protein in *Saccharomyces cerevisiae*.

To identify additional proteins in the mating-type vegetative incompatibility pathway, TOL was used as bait in a yeast two-hybrid system. One of the putative *pit* genes (<u>p</u>roteins that <u>i</u>nteract with <u>T</u>OL) was *ncvip1*, which has homology with *vip1*, an uncharacterized gene in *Schizosaccharomyces pombe*. RIP (<u>r</u>epeat-<u>i</u>nduced <u>p</u>oint) mutational analyses of the *ncvip1* was conducted and progeny examined as to their effects on mating-type associated vegetative incompatibility. At the termination of this study, the influence of *ncvip1* remained unclear. Strains were constructed that will provide the tools for future analyses.

A novel mutation was discovered that caused a partial <u>sup</u>pression of <u>m</u>ating-type <u>v</u>egetative <u>i</u>ncompatibility (Sup [mvi]) phenotype. This mutation appears to be linked to or in *mat a*, localizing the gene causing the Sup (mvi) phenotype to linkage group I. Partial diploid analysis and complementation assays failed to determine if the mutation was in *mat a* or not and demonstrated that the Sup (mvi) phenotype does not suppress mating-type vegetative incompatibility when the mating-type gene products are in the same nucleus. Further analysis is needed to determine the nature of the mutation and the full influence of the Sup (mvi) phenotype in a homoallelic condition.

TABLE OF CONTENTS

TABLES

FIGURES

ACKNOWLEDGMENTS

The process of completing a Masters degree is no minor undertaking and I've managed to complete it with my sanity intact. In the midst of the bench work and writing, I have learned to love working with *Neurospora* and trips in the field have given me a once in a lifetime chance to participate in ecological aspects of *Neurospora*. How else would I have been able to see so many part of the United States. Thanks Dave!

I've relied on my family and friends for support to me through the frustrations of research and have asked numerous questions of my colleagues. I thank all of you from the bottom of my heart for your support. You've all been amazing beyond words. I also want to make a dedication to my friend Jamie, who I lost in the spring of 2000. You will never be forgotten.

Finally, I want to thank all of the past and present members of my lab for providing endless insight and knowledge. In this thanks, I would like to include all the feedback that may not have been the most desired, but is ultimately the most critical to good research. Louise also deserves recognition for personal encouragement when I needed it most. I've also been fortunate to have had financial support from NSERC, Louise and UC Berkeley. I've been given many educational and personal learning opportunities in my exposure to the Universities of British Columbia and the California, Berkeley, which I appreciate greatly. Who knew that this is where I would end up or that I would have had the chance to learn from so many great people. Thank-you for everything.

1. INTRODUCTION AND BACKGROUND

1.1 THESIS OBJECTIVES

Neurospora crassa is an important model system due to its longstanding use in the research community for general biology, genetics and biochemical analyses. The recent completion of the *N. crassa* genome sequencing project lends *N. crassa* the comparative edge needed to compete in the new age of research; genomic analysis and proteomics will help answer global regulation and biological questions that were impossible to investigate even ten years ago.

The goal of this project is to investigate the phenomenon of vegetative incompatibility in *N. crassa*, which serves as a self/non-self recognition system in filamentous fungi. The complete function and purpose of this phenomenon is not clearly understood, but remains a ubiquitous process among filamentous fungi. The mating-type (*mat*) locus is a known heterokaryon incompatibility (*het*) locus in *N. crassa* and is the root of the pathway investigated in this project. Genetic differences at the mating-type locus between strains undergoing fusion in the vegetative state, render a physical incompatibility reaction in which the fusion cell undergoes a cell death process. In this particular pathway, a separate locus, *tol*, is known to mediate the response of MAT a-1 and MAT A-1 in mating-type associated vegetative incompatibility; if both strains are *tol-*, the heterokaryon is compatible. The extent of *tol*'s role in this process is poorly understood. Identification and characterization of clones that encode TOL-interacting proteins were the focus of this research. Therefore, the primary goals of this project are:

I. To identify regions of similarity and possible important domains of TOL within a number of TOL-like sequences identified in the recently completed genome sequence.

II. To identify additional components in the mating-type associated vegetative incompatibility pathway, with the possibility of elucidating additional information about the general process of vegetative incompatibility.

1.2 THE LIFE CYCLE OF *Neurospora crassa*

Neurospora crassa Shear and Dodge is an excellent eukaryotic source of genetic and molecular information; it is easy to grow and manipulate in a laboratory environment, has a reasonable generation time, and there are a number of loci with known mapped mutations, leading to its development as a model organism for research in eukaryotic systems. An example of its use is in the studies of Beadle and Tatum, who used *N. crassa* to deduce the "one gene, one enzyme" theory that won them the Nobel Prize in Physiology and Medicine in 1958 (Davis, 2000).

Neurospora crassa was first described in 1927 by Shear and Dodge as a haploid, heterothallic organism, indicating it requires two mating types (designated *mat A* and *mat a*) to complete its life cycle (Shear and Dodge, 1927). Once thought to be a semi-tropical to tropical fungus, it has recently been found in a wide geographical range, while maintaining the requirement of an intense vegetation burn (D. Jacobson, D. Natvig, J. Taylor and L. Glass, personal communications; Jacobson *et al.*, 2001). The life cycle of *N. crassa* (**Figure 1**) begins with ascospore germination in response to heat activation, which permits the proliferation of the vegetative form of the fungus, the filamentous homokaryotic hyphae. *N. crassa* hyphae are partially septate, multinucleate, tube-like cells. These hyphae grow by tip extension and hyphal fusion, resulting in a complex network within or between *N. crassa* strains. In order for the organism to expand into new niches, new colonies form through re-establishment of hyphal fragments or via spores, such as those produced by modified hyphal structures called conidiophores (micro- and macro- conidia - the asexual spores) or sexual ascospores. Conidia are produced during the vegetative life cycle and also function as male gametes in the sexual reproductive cycle (Springer, 1993).

Figure 1. The life cycle of *Neurospora crassa*
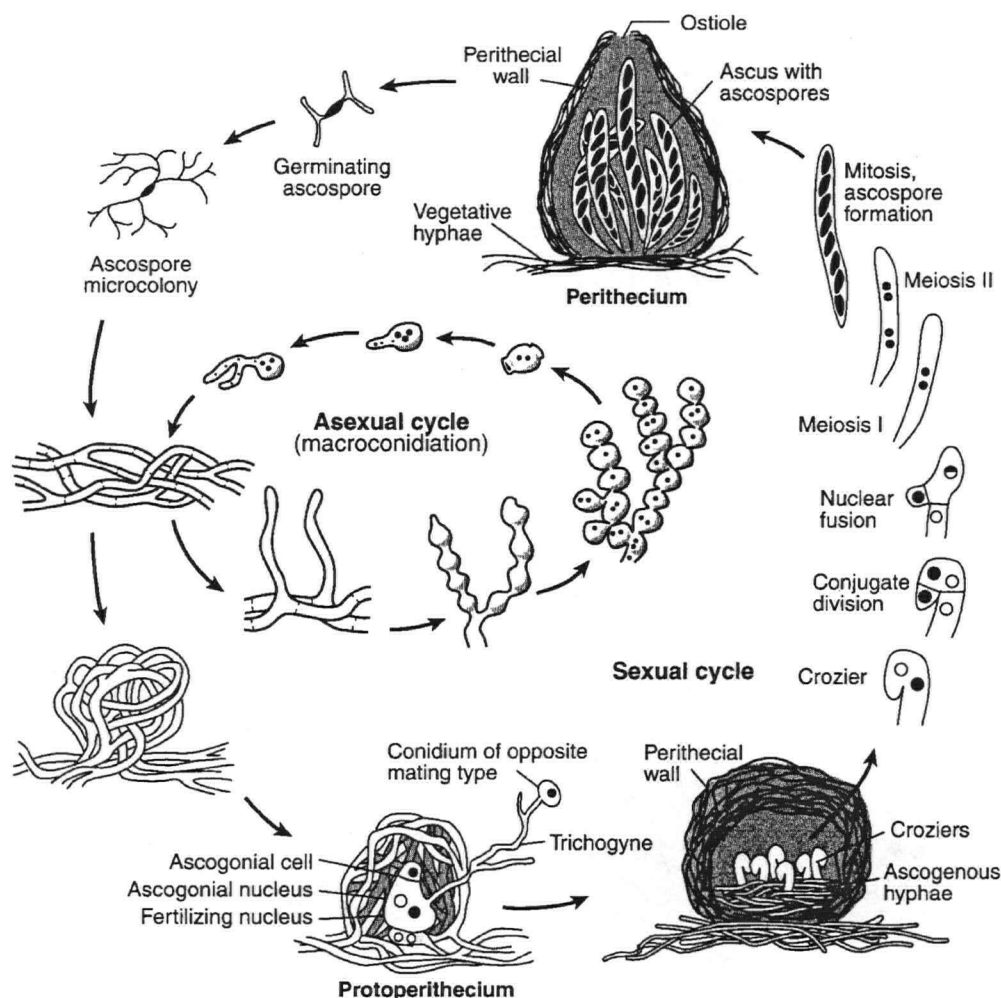
Figure 1. The basic life cycle of *N. crassa* (taken from Davis, 2000). See text for details.

N. crassa cultures are either *mat A* or *mat a* mating-type, and fusion of opposite mating-types is required for completion of the sexual cycle. Both mating-types are capable of forming the male and female structures necessary for sexual development. Coiled female structures, protoperithecia, form in response to nutrient deprivation (i.e. nitrogen starvation – Westergaard and Mitchell, 1947; Johnson, 1978). Pheromone signaling is proposed for the initial recognition

of the presence of a strain of the opposite mating-type. There is subsequent growth of a trichogyne (a modified female outgrowth) towards the male partner of the opposite mating-type (Bistis, 1981; pheromone precursor genes – D. Bell-Pederson and D. Ebbole, unpublished results; reviewed in Shiu and Glass, 2000). Male structures - microconidia, macroconidia or hyphae - fuse with the trichogyne, resulting in plasmogamy and the subsequent initiation of the sexual cycle (Raju, 1980; Raju, 1992). After plasmogamy, the male nucleus travels down the trichogyne to pair with a genetically different female nucleus, resulting in a *mat A* : *mat a* nuclear pair. A proliferation of dikaryotic (ascogenous – N+N) hyphae forms and the surrounding haploid hyphae grow to generate a protective structure resulting in the perithecium. The formation of a crozier (hook-like in appearance) in the ascogenous hyphae partitions a pair of opposite mating-type nuclei and permits a synchronous mitotic division. Karyogamy takes place in the penultimate cell of the crozier and forms the ascus, a sac-like structure characteristic of ascomycota. The paired N+N nuclei fuse in the crozier compartment and the diploid cell undergoes meiosis followed by mitosis to produce an ordered tetrad of eight haploid ascospores (Raju, 1980), four of each mating-type. Approximately 200 to 400 asci typically result from one fertilization event.

1.3 HYPHAL ANASTOMOSIS

Vegetative cell growth occurs through hyphal tip extension and hyphal tip fusions (anastomoses) within and between filamentous fungal individuals, permitting the formation of complex networks in the natural environment. Anastomosis is poorly understood in many fungi, but has an essential role in how filamentous fungi interact on an individual and organismal level.

4

As described below, anastomosis is a key step in vegetative incompatibility. When hyphae detect nearby hyphae, there are negative and positive autotropism response signals (Bhuiyan and Arai, 1993; Garnjobst and Wilson, 1956; Gooday, 1975). Initiation of hyphal fusions and pre-contact may be mediated by chemical signals, which determine whether or not the hyphae are to undergo a fusion process. This process involves the physical contact of the opposing hyphae, breakdown of the cell wall and the subsequent fusion of the plasma membranes (Buller, 1933; Glass *et al.*, 2000; Hickey *et al.*, 2001). After fusion, there is an intermixing of cytoplasmic contents and often an altering in cytoplasmic flow. The consequences of these shifts are currently unclear. The characteristics described have been elegantly demonstrated in *in vivo* using confocal fluorescence imaging and a variety of specific dyes in wild type *N. crassa* strains and other filamentous fungi (Fisher-Parton *et al.*, 2000; Hickey *et al.*, 2001; http://www.neurospora.com). Recently identified mutants, *ham-1* and *-2* (*h*yphal *a*nastomosis *m*utant; Wilson and Dempsey, 1999; Q. Xiang and L. Glass, unpublished results), *mak-2* (*ma*p *k*inase - FUS3 homologue; Bobrowicz and Ebbole, 2001) and *nrc-1* (*n*on-*r*epressible *c*onidiation - STE11 homologue; Kothe and Free, 1998), are defective in hyphal fusions in *N. crassa*, both with other individuals and with themselves. These mutants are currently being used to further investigate the overall process of hyphal anastomosis.

1.4 VEGETATIVE INCOMPATIBILITY

Vegetative (or heterokaryon) incompatibility is the phenotype generated from the fusion of hyphae between individuals with one or more dissimilar heterokaryon incompatibility (*het*) alleles (or *vic* for *v*egetative *inc*ompatibility) at one or more loci (**Figure 2**). This is a

phenomenon observed in many filamentous fungi, in both ascomycota and basidiomycota (Beadle and Coonradt, 1944; Glass and Kuldau, 1992).

Figure 2. Vegetative growth and incompatibility
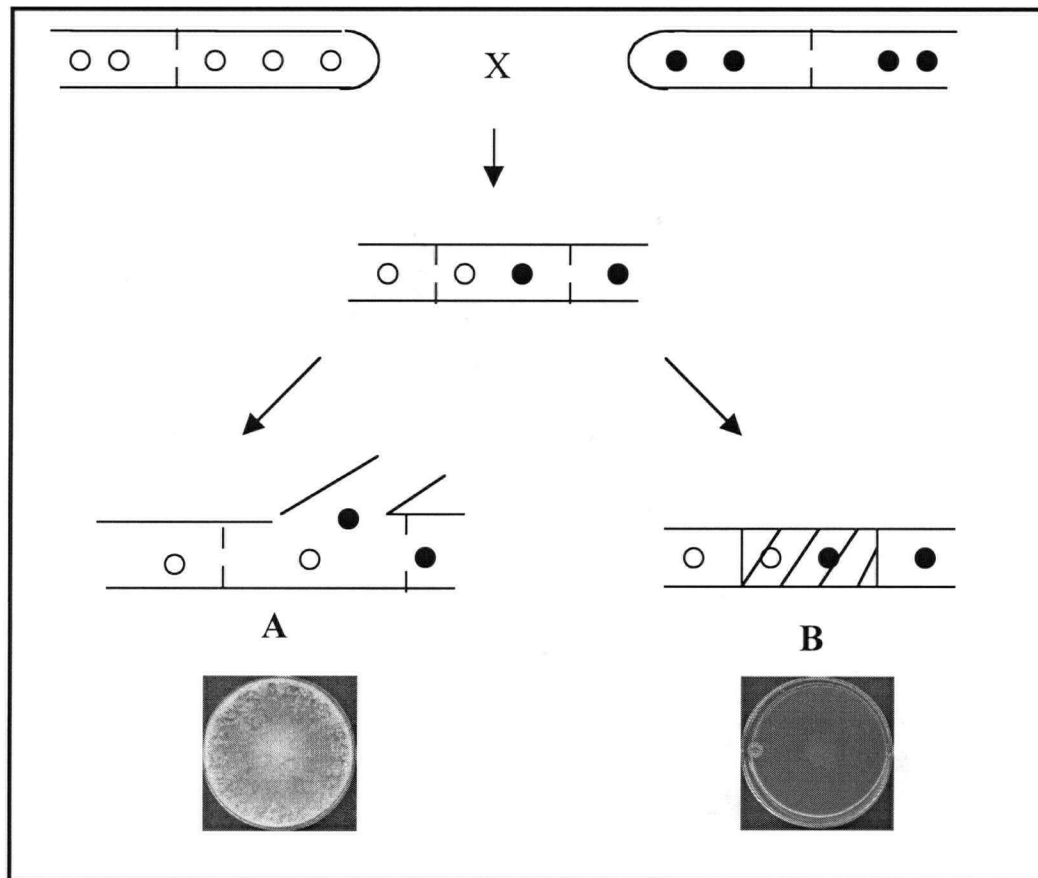
Figure 2. When two filamentous fungal hyphae with genetically different nuclei (open and solid circles) fuse during vegetative growth, there are two possible outcomes. Either the two nuclei are compatible and can continue normal growth as a heterokaryon (**A**) or the cell fusion results in <u>vegetative incompatibility</u>, which is demonstrated through inhibited cell growth, aconidiation and/or an increase in the percentage of cell death (**B**). Vegetative incompatibility is also characterized with septal plugging and cell death in the heterokaryon cell (hatching). het test plates (Vogel's minimal medium; Vogel, 1964): **A** = FGSC 5448 + I-20-26 (*ad-3A; inl A + ad-3B arg-1 A* = compatible); **B** = FGSC 7214 + R1.21 (*ad-8 lys-5 A + pyr-4 arg-5 pregc-2 trp-3 a* = incompatible).

1.4.1 Relevance of vegetative incompatibility

het loci are believed to serve as a self/non-self recognition system in filamentous fungi, demonstrated through the recognition of protein products from alternative alleles present in the same cytoplasm and the subsequent vegetative incompatibility response (Glass and Kuldau, 1992; Glass *et al.*, 2000). Despite the awareness that this serves as a self/non-self recognition system, the significance of vegetative incompatibility is not fully understood. There is recognition of the presence of different polypeptides from these loci, indicating a role similar to that of an immune system, but present in filamentous fungi (Caten, 1972). An alternative theory states that the rise in polymorphisms at *het* loci is an accident and that there is no biological relevance to vegetative incompatibility (Bégueret *et al.*, 1994).

To support the first argument of a self/non-self recognition system, it was found that *het-c*, an allelic *het* locus in *N. crassa*, is under balancing selection, a characteristic of other self/non-self systems such as the *S* locus in plants and the MHC in humans (Wu *et al.*, 1998). Maintenance of the specificity region in *het-c* favors grouping by allele type and not species, a prevalence of nonsynonymous over synonymous substitutions and demonstrates equal allele frequencies in a population and trans-species polymorphisms (Wu *et al.*, 1998). These are common characteristics of balancing selections. Selection of *het-c* as a self/non-self regulating locus is also supported by its existence in multiple species with the Sordariaceae, spanning approximately 35 million years (Wu *et al.*, 1998). Transformation of *het-c* into *Podospora anserina* and subsequent alteration in growth and morphology in transformants indicate maintenance of function of *het-c* (Saupe *et al.*, 2000). However, the *het-c* homologue in *Podospora anserina* (*hch*), which has approximately 75 million years divergence, is apparently not polymorphic (Saupe *et al.*, 2000). A more recent study examining the mating-type gene, *b1*,

7

in the basidiomycete *Coprinus cinereus* supported a theory of balancing selection for maintenance of polymorphisms at mating-type loci (May *et al.*, 1999). A study looking at six *vic* (vegetative incompatibility) loci in European and North American isolates of *Cryphonectria parasitica* did not completely support the balancing selection theory for all het loci. Most allele frequencies varied from the expected value of 0.5 and mating did not appear to be random among the European isolates (Milgroom and Cortesi, 1999). This may indicate that not all *het* loci are under balancing selection.

Another likely function of vegetative incompatibility is the reduction in genetically different strains forming viable heterokaryons; heterokaryons are rarely isolated from nature (Mylyk, 1976; supported in simulation studies by Pandit and Maheshwari, 1996). The reduction in viable heterokaryon formation is due to the large number of segregating *het* loci, which decreases the likelihood of mycoviral transfers or the transfer of deleterious elements through a mycelial community. This has been demonstrated in Aspergillus with viral transfer (van Diepeningen *et al.*, 1997 and 1998), cytoplasmic elements in Neurospora (Caten, 1972), and plasmid transfer of virus-like dsRNA in *N. crassa* (Debets *et al.*, 1994). Prevention of resource plundering is one of the latest proposals for a function of vegetative incompatibility, since pre-established colonies will prefer to maintain their priority over local resources (Debets and Griffiths, 1998).

1.4.2 Characteristics of vegetative incompatibility and programmed cell death

Inhibited growth, cell death and lack of conidiation in heterokaryons characterize vegetative incompatibility in filamentous fungi. In *N. crassa*, the vegetative incompatibility

response causes the affected cell to undergo septal wall plugging, cytoplasmic shrinkage from the plasma membrane, and the re-packaging of cell contents into smaller bodies for degradation (Jacobson *et al.*, 1998). Biological examination of the cell death response has shown a pattern that is distinct from necrosis (Jacobson *et al.*, 1998; Garnjobst and Wilson, 1956; Caten, 1972). The pattern of programmed cell death in vegetative incompatibility is currently being investigated, with respect to *het-c* in *N. crassa*, to determine if biochemical similarities exist between vegetative incompatibility and programmed cell death in higher eukaryotes (Marek *et al.*, 1998). Additional characteristics, such as DNA degradation in *N. crassa* (assessed via TUNEL [terminal deoxynucleotidyl transferase-mediated d<u>U</u>TP-X <u>n</u>ick <u>e</u>nd <u>l</u>abeling] assay; Marek *et al.*, 1998) and reduction in RNA production and new protein synthesis in *P. anserina* (Labarére *et al.*, 1974; Boucherie *et al.*, 1981) are also observed in incompatible heterokaryons. These attributes, as well as presence of DNA laddering from DNA degradation and the presence of caspases typify apoptosis (Saraste and Pulkki, 2000); *N. crassa* appears to lack DNA laddering and caspases. Database searches with caspases and other proteins from known apoptotic components with the *N. crassa* genome revealed very few conserved constituents (S. Sarkar, A. Pandey and L. Glass, unpublished results). Because the cell death of vegetative incompatibility is still currently under investigation and does not have all the components of apoptosis, the term programmed cell death is preferred.

Cell death in *N. crassa* is assessed by the percentage of dead cells or hyphal compartments in a sample. In incompatible heterokaryons, there is typically an increase in cell death compared to the low cell percentage of cell death seen in individual strains or compatible heterokaryons. Approximately 15-30% cell death is observed in partial diploids (explained in section **3b2.1.5.1**) or *het-c* or *het-6* incompatible transformants (Jacobson *et al.*, 1998; L. Glass,

9

unpublished results). Evan's Blue is a vital dye used to detect cell death by differentiating live and dead tissue via its exclusion from live cells, permitting the identification of dead compartments with microscopy (Gaff and Okong'O-Ogola, 1971).

1.4.3 Vegetative incompatibility loci in filamentous fungi

Molecular characterization of vegetative incompatibility loci is underway for *N. crassa*, *P. anserina, A. nidulans* and *C. parasitica* (**Figure 3**). There are two types of vegetative incompatibility systems in filamentous fungi – allelic and non-allelic. Of the identified *het* loci in *N. crassa*, most are allelic; allelic incompatibility results when there is a genetic difference between strains differing at only one *het* locus. *un-24* and *het-6* are the first putative set of non-allelic incompatibility loci in *N. crassa* (Smith *et al.*, 2000b), where incompatibility is caused by an interaction between two different *het* loci. However, because of the close linkage of *un-24* and *het-6*, in addition to chromosomal rearrangements, these two loci segregate as a single locus in a cross (M. Smith, personal communications). *het-c* and mating-type, two other well-characterized *het* loci, are allelic.

Figure 3. Vegetative incompatibility loci in filamentous fungi

N. crassa	11 *het* loci ; 10 allelic and one non-allelic (Garnjobst and Wilson, 1956; Perkins, 1975; Mylyk, 1975; Glass and Kuldau, 1992; Saupe, 2000)
	Cloned and characterized:
	    het-c : 3 natural alleles (Saupe *et al.*, 1996b and 1997; Wu *et al.*, 1998; Wu and Glass, 2001)
	    Mating-type : 2 idiomorphs (Glass *et al.*, 1988; Glass *et al.*, 1990; Staben and Yanofsky, 1990; Shiu and Glass, 2000)
	    het-6 : non-allelic with *un-24* (Smith *et al.*, 1996, 2000a and 2000b)

P. anserina	9 *het* loci ; 5 allelic (*het-b*, *het-q*, *het-r*, *het-s*, and *het-z*) and 4 non-allelic (Bégueret *et al.*, 1994; Saupe, 2000)
	3 cloned and characterized
	    het-s : allelic; prion analogue (Coustou *et al.*, 1997)
	        het-c : non-allelic with *het-e* and *het-d* (Saupe *et al.*, 1995)
	    het-e : non-allelic with *het-c* (Saupe *et al.*, 1994)

A. nidulans	8 *het* loci; allelic (Dales and Croft, 1983; Anwar *et al.*, 1993)
C. parasiticus	6 *het* loci; allelic (Milgroom and Cortesi, 1998)
General reviews:	Bégueret *et al.*, 1994; Nelson, 1996; Glass *et al.*, 2000; Saupe, 2000

Figure 3. A literary description of vegetative incompatibility loci in a number of filamentous fungi, including references.

Eleven identified heterokaryon incompatibility (*het*) loci regulate vegetative incompatibility in *N. crassa* (**Figure 4**); for normal vegetative growth, fusing hyphae must be isogenic at all *het* loci. Mating-type, *het-c* and *het-6* loci have been mapped, cloned and characterized in *N. crassa* (Glass *et al.*, 1990; Staben and Yanofsky, 1990; Saupe *et al.*, 1996b; Smith *et al.*, 1996 and 2000b).

Figure 4. Vegtative incompatibility and associated loci in *N. crassa*
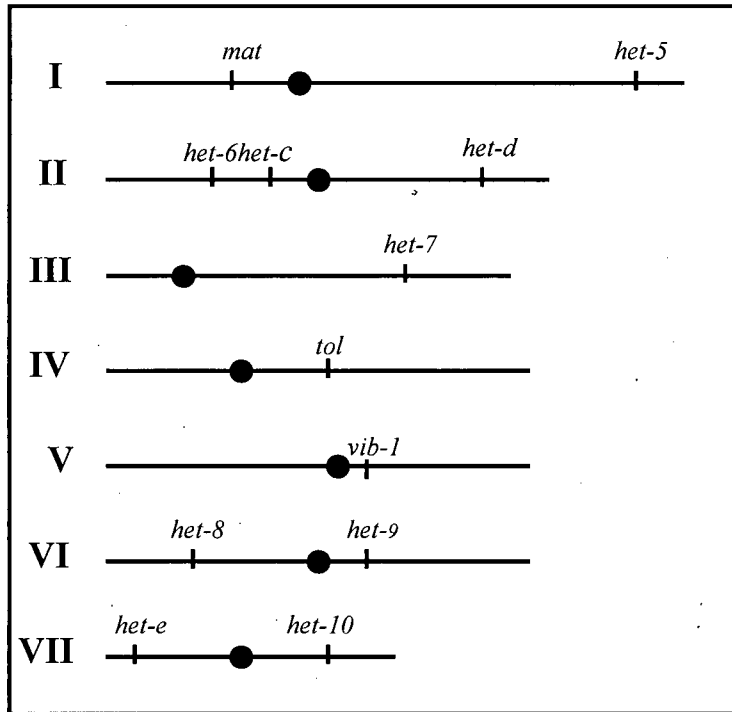
Figure 4. Heterokaryon incompatibility (*het*) loci in *N. crassa* and other related genes (modified from Perkins *et al.*, 2001) and their corresponding linkage groups (I through VII). *mat* = mating-type; *het* = heterokaryon incompatibility; *tol = tolerance; vib-1* = vegetative incompatibility blocked. *het-i* is not included on this map, but is proposed to be located on either linkage groups I or II (Perkins *et al.*, 2001). Centromeres are represented by filled circles, horizontal lines represent the linkage groups, and vertical lines represent the genes.

Alternative functions for the gene products of *het* loci are postulated due to their existence in many fungal species. Mating-type in *N. crassa* is required for the entry and completion of the sexual cycle; *het-6*'s putative non-allelic partner, *un-24*, has homology with ribonucleotide transferases in a number of organisms (Smith *et al.*, 2000a and b). *het-c* does not have an apparent alternative function in addition to vegetative incompatibility (Saupe *et al.*, 1996b).

1.4.3.1 A model vegetative incompatibility locus: *het-c* in *N. crassa*

het-c was first identified in 1953 with heterokaryon tests by Garnjobst (Garnjobst, 1953) and later via partial diploid analyses (Perkins, 1975; partial diploid analysis explained in section

3b.2.1.5.1). The phenotype of *het-c* vegetative incompatibility is not as severe as mating-type or *het-6*, but characteristic – aconidial, slow growth rate with curly hyphal edges, and an increase in cell death as assayed with Evan's Blue staining. *het-c* was cloned and characterized and encodes a 966 amino acid polypeptide (Saupe *et al.*, 1996b). There are three identified natural alleles of *het-c* - Oak Ridge (OR), Groveland (GR) and Panama (PA); each has specific insertion/deletion (in/del) pattern in a characterized specificity domain (amino acids 247-284 in the Oak Ridge allele – Wu *et al.*, 1998; Wu and Glass, 2001). If the specificity domain is switched between alleles, the transformed strain will adopt the identity of the newly inserted fragment, as demonstrated by transformation assays and heterokaryon tests (Saupe and Glass, 1997; Wu *et al.*, 1998; Wu and Glass, 2001).

het-c is subject to balancing selection and population analyses have been completed using *het-c* to demonstrate that *het-c* has characteristics of a self/non-self recognition system (Wu *et al.*, 1998; section **1.4.1**). Ultrastructural studies of *het-c* vegetative incompatibility have also been conducted, demonstrating characteristics of a programmed cell death process (Jacobson *et al.*, 1998; Marek *et al.*, 1998; section **1.4.2**). The latest application of *het-c* in investigating vegetative incompatibility is using HA and GFP tagged HET-C proteins to detect heterocomplex formation; heterocomplex formation between alternative HET-C proteins apparently triggers vegetative incompatibility. Co-immunoprecipitation with GFP and HA tagged combinations of the three different HET-C polypeptides identifies a membrane fraction complex containing alternative polypeptides only in incompatible heterokaryons (G. Iyer and L. Glass, unpublished results). This observation has been supported using indirect immunofluorescence confocal microscopy to identify the location of tagged HET-C in incompatible transformants (S. Sarkar and L. Glass, unpublished results). In both experiments, HET-C was undetectable in control

compatible transformants or individual wild type strains transformed with the tagged *het-c* constructs. These results support the theory that heterocomplex formation between alternative HET-C proteins is integral to the process of vegetative incompatibility and is normally localized to the plasma membrane in an incompatible response.

1.4.3.2 *N. crassa*: *het-6* and *un-24*

het-6, with two identified alternative alleles (Oak Ridge [OR] and Panama [PA]), was originally identified as poorly growing strains from a partial duplication cross involving linkage group II (Mylyk, 1975). In 1996, Smith *et al.* (1996) examined a partial diploid escape strain involving the *het-6* locus to determine the nature of the resulting vegetative compatibility. Transformations of *het-6*^{PA} into a *het-6*^{OR} strain were compatible (versus the expected incompatible), leading to the hypothesis that *het-6* may be involved in a non-allelic incompatible interaction; the non-allelic partner would have to be linked to *het-6* to cause vegetative incompatibility in a partial diploid. Transformation reduction assays showed two non-overlapping fragments from a genomic cosmid caused vegetative incompatibility, one of which contained *het-6*. It was proposed that *un-24* was the non-allelic partner of *het-6* since no recombination with *het-6* was detected in 220 progeny and *un-24* was located on the other genomic fragment that conferred vegetative incompatibility. This demonstrated the first set of non-allelic vegetative incompatibility genes in *N. crassa* (Smith *et al.*, 2000a and b).

het-6/un-24 have a severe vegetative incompatibility phenotype, including a slow growth rate (Mylyk, 1975; Smith *et al.*, 1996), aconidiation and an increase in cell death (Jacobson *et al.*,

1998). There is a lack of recombination between *het-6* and *un-24* due to an inversion between the *het-6^{OR}* and *het-6^{PA}* alleles (Smith, personal communications).

The *un-24* mutant was identified in crosses between translocation strain, T(II→III)*AR18*, and wild type strains. This permitted the generation of partial diploid progeny in the region containing *un-24*; RIP (**r**epeat-**i**nduced **p**oint) mutations occur in regions with duplicated sequence and results in G:C→A:T transitions in both copies of the duplicated region (section **3a.2.2.6.1**). One of the RIP progeny was the temperature sensitive *un-24* strain DJ949-90, which grows normally below 30°C, but requires high osmolarity conditions above 34°C for normal growth. The vegetative incompatibility function appears to be independent of temperature sensitivity (Smith *et al.*, 2000b). UN-24 has homology with a class I ribonucleotide reductases, which transform nucleoside diphosphates (NDPs) into deoxynucleoside diphosphates (dNTPs) and are essential for DNA synthesis (Smith *et al.*, 2000a). The mutation in *un-24* DJ949-90 may alter the protein function in a key binding region for ATP/dATP, thus abolishing normal function and folding of the large subunit of the ribonucleotide reductase in *N. crassa* above 34°C.

het-6 has been cloned and shows no similarity with known proteins. It does, however, contain three regions, collectively called the SET domain, that have similarity with HET-E in *P. anserina* and TOL in *N. crassa* (SET domain name for HET-**S**IX, HET-**E** and **T**OL; Smith *et al.*, 2000b; Saupe *et al.*, 2000 - **Figure 5**). HET-6, HET-E and TOL are involved in vegetative incompatibility pathways and the domains of similarity may be key in the role they play in the overall pathway.

Figure 5. SET domain alignments

```
NcHET-6PA     53  VP SQA SY ALSYVWGDSTRTH I VVN  NDGRGAFVTLRLTTSLDTCLR  104
NcHET-6OR     53  AP SPPPSY ALSYVWGDSTRTH I VANE NDGR-AFIPLRLTSSLDTCLR  103
PaHET-E       18  LPSGKIPPYAILSHTWGPDE---EVSYK KDGR-AVSKLGYN-K RFCAD    64
NcTOL        336  LA SETP-FMSLSHCWGKDG---VP QLLKGNYDRFTKEG RLT-ELPKTFR 382


NcHET-6PA    117  PLPLWIDQ CINQ DDAEKSSQ LLMKN YSSA  149
NcHET-6OR    116  PLPLWIDQ CINQ DNEEKSFQ RLMRD YSSA  148
PaHET-E       71  RKFFW DTCCIDKSNSTELQEA NSMFRWYRDA  103
NcTOL        395  P-YTWIDS CIIQ SKE WDDESVKMQY YRNS  426


NcHET-6PA    234  WFTRVWT QEFCLCSDTVF  252
NcHET-6OR    233  WFKRLWT QEFCLCADTIF  251
PaHET-E      135  WFTRGWT QELIAPTSVEF  153
NcTOL        491  LFTRGWV QEQLLARRTLI  509
```

Figure 5. The regions in common in the **SET** domain: HET-**S**IX Oak Ridge (OR) and Panama (PA) alleles from *N. crassa*, HET-**E** from *P. anserina*, and **T**OL from *N. crassa*. Identical residues are shown in black blocks with white font, similar residues are shown in grey blocks with white font. Alignment taken from Saupe *et al.* (2000) and using the CLUSTALW (Thompson *et al.*, 1994) program and BOXSHADE server (http://www.ch.embnet.org/software/BOX_form.html).

1.4.3.3 *Podospora anserina: het-c/het-e*

Non-allelic incompatibility results in an incompatibility response between polypeptides at two or more unlinked loci. In *P. anserina*, there is a key distinction between allelic vegetative incompatibility and non-allelic incompatibility, where allelic loci also cause sexual incompatibility. *het-c/het-e* non-allelic incompatibility is noted by an initial heterokaryon formation and/or growth from a germinated ascospore, followed by a growth arrest and severe lytic reaction that involves vesicle formation and destruction of cellular compartments.

het-e is a component of one of the three characterized sets of non-allelic vegetative incompatibility loci in *P. anserina*; *het-e* has four different alleles, as does its non-allelic partner

het-c (not to be confused with *het-c* in *N. crassa*). *het-c* in *P. anserina* has an additional non-allelic partner, *het-d*. For the purpose of this study, focus will be made regarding the *het-c/het-e* interactions. *het-e*1 encodes a 1356 amino acid polypeptide, with a GTP binding domain and WD40 repeat region (Espagne *et al.*, 1997). WD40 repeats are found in a large variety of proteins in different organisms with a number of different functions - from signal transduction and gene regulation, to vesicular trafficking and cytoskeleton assembly (Espagne *et al.*, 1997). The N-terminal region of HET-E has a putative GTP-binding region that is crucial for function; a mutation from Lys→Asp at amino acid position 306 in the region encoding a putative GTP-binding domain abolishes vegetative incompatibility. The C-terminal portion has a repeat of 42 amino acids that is homologous to the β-subunits of the heterotrimeric G-proteins. These G proteins are involved in protein-protein interactions and the region in HET-E1 seems to be functional and necessary for vegetative incompatibility (Espagne *et al.*, 1997). The number of WD40 repeats varies from three to twelve in different *het-e* alleles, but is not essential for allelic specificity (Espagne *et al.*, 1997). Mutational analysis of *het-e* by gene replacement did not reveal any additional phenotypes. Of interest to this study, HET-E also contains a region with three domains similar to HET-6 and TOL from *N. crassa* (Smith *et al.*, 2000b; Saupe *et al.*, 2000; **Figure 5**). The significance of this region in unknown, but it is found in proteins known to be involved in vegetative incompatibility and therefore may represent a novel domain in this particular pathway or pathways.

het-c, in *P. anserina*, has at least four different alleles and encodes a 208 amino acid polypeptide, with some sequence similarity with a glycolipid transfer protein from pig brain (Saupe *et al.*, 1994). Homozygous crosses between *het-c* mutants disrupt the maturation of normal ascospores, therefore indicating function of HET-C in addition to vegetative

incompatibility. There is also a region of variability between the four *het-c* alleles that have been cloned (position 126-160). The variability in *het-c* shows characteristics of being under selection to permit allelic specificity in the vegetative incompatibility response (Saupe *et al.*, 1995). This correlates with evidence in *het-c* in *N. crassa*, which appears to be under balancing selection (section **1.4.1**).

mod-A (*mod*ifier) suppresses non-allelic vegetative incompatibility, although complete suppression of *het-c/het-e* incompatibility must be accompanied by an additional mutation in the unlinked locus, *mod-B*. However, the double mutant, *mod-A1 mod-B1*, has conditional female sterility, depending on the medium, age of the mycelia, or temperature (Boucherie and Bernet, 1980). *mod-A* encodes a 687 amino acid polypeptide, with a C terminal region that is proline rich and resembles a SH3-binding domain (Anwar *et al.*, 1993).

1.4.3.4 *N. crassa:* mating-type

Successful completion of the sexual phase requires plasmogamy and karyogamy of *mat a* and *mat A* nuclei in the ascus (Beadle and Coonradt, 1944; Garnjobst, 1953). In contrast to the vegetative phase, the mating-type locus does not appear to function as an incompatibility locus during the sexual cycle, allowing co-existence of *mat A* and *mat a* nuclei in the same hyphae. The apparent absence of signals conferring vegetative incompatibility during the sexual cycle is also true of the other *het* loci.

The mating-type locus in *N. crassa* serves as a vegetative incompatibility locus, as well as being an integral factor of the sexual cycle. Not all filamentous fungi have this dual role of the mating-type locus, but *Ascobolus stercorarius* (Bistis, 1994), *A. heterothallicus* (Kwon and

Raper, 1967) and *Sordaria brevicollis* (J. Bond, personal communications) have mating-type associated vegetative incompatibility. The mating-type locus in *N. crassa* is present as one of two co-dominant "alleles" – *mat A* and *mat a*, which map to the left arm of linkage group I (**Figure 4**). The sequence composition of each "allele" is very different despite the virtually identical flanking regions (Glass *et al.*, 1988), so the term idiomorph has been adopted (Metzenberg and Glass, 1990). The term idiomorph demonstrates that the two alternative sequences at the *mat* locus are not structurally related or related by recent common decent.

The *mat A* and *mat a* idiomorphs have been cloned and characterized (Glass *et al.*, 1990; Staben and Yanofsky, 1990; **Figure 6**). The *mat A* idiomorph is 5301 bp and has three characterized open reading frames (ORFs) – *mat A-1*, *mat A-2* and *mat A-3*. *mat A-1* (1.5 kb) confers vegetative incompatibility associated with mating-type and mating-type identity, via a 293 amino acid polypeptide (Glass *et al.*, 1990). Null mutants of *mat A-1* permit heterokaryon formation with *mat A* and *mat a* strains, but are sterile (Griffiths, 1982; Saupe *et al.*, 1996a), although one mutant, A^{m99}, has no vegetative incompatibility and retains partial female fertility but can not complete a cross as a male (Saupe *et al.*, 1996a). A mutation in A^{m99} converts a tryptophan residue at position 86 to a stop codon. Two other deletion constructs, A^{m54} and A^{m64}, are sterile and retain incompatibility function upon transformation into a *mat a* strain, but not in a heterokaryon (Saupe *et al.*, 1996a). A region of MAT A-1 shows similarity to the DNA binding motif MATα1 in *S. cerevisiae*, a mating-type transcriptional regulator (Glass *et al.*, 1990). This homology is conserved in many *mat A-1* equivalents in other fungi (Shiu and Glass, 2000).

Figure 6. Mating-type idiomorphs in *N. crassa*
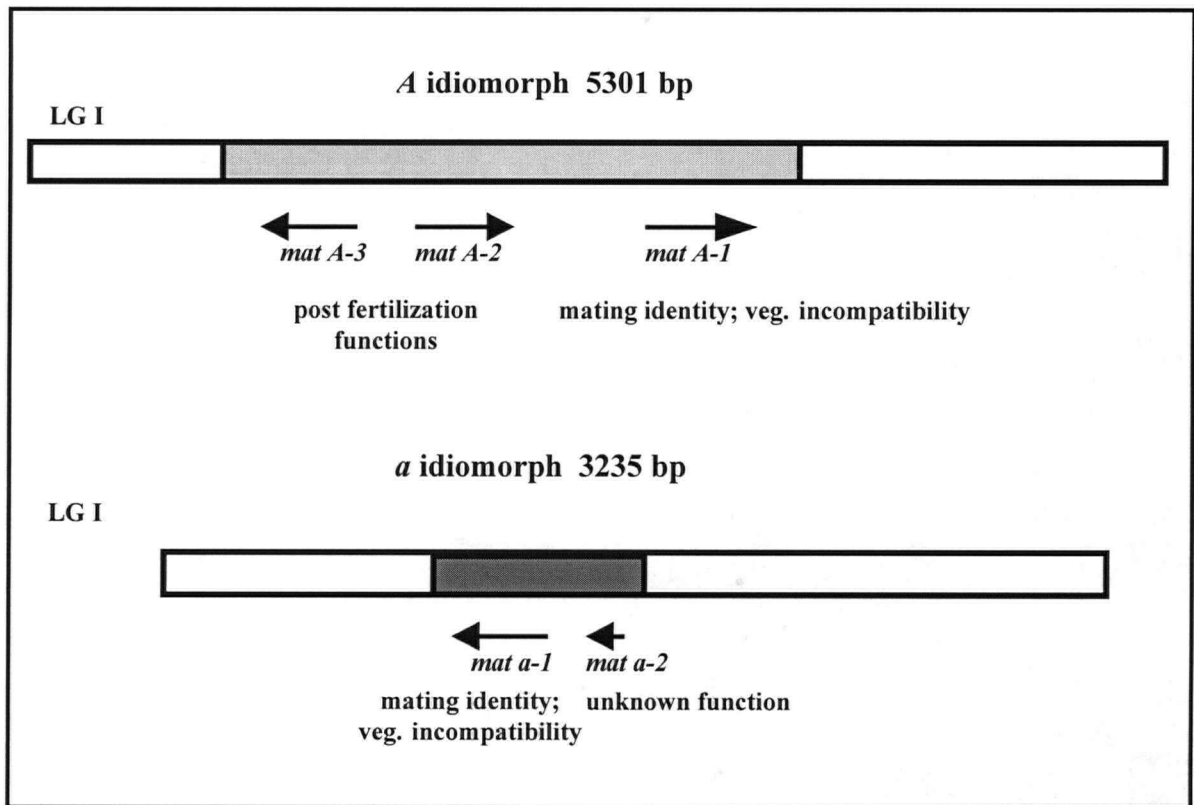
Figure 6. The mating-type locus of *N. crassa*, found on linkage group I (modified from Glass and Nelson, 1994); *mat A* and *mat a* idiomorphs. Additional *mat a-2* putative open reading frame contributed by Pöggeler and Kück (2000). See text regarding details about the ORFs of each idiomorph.

mat A-2 and *mat A-3* are 1.6 kb and 2.4 kb respectively; both identified through RIP analysis (Ferreira *et al.*, 1996 and 1998). MAT A-2 resembles SMR1 from *P. anserina*, a gene encoded in *mat⁻* and required for post-fertilization function (Debuchy *et al.*, 1993). Mutations in *mat A-2* have no vegetative phenotype, but appear to affect post-fertilization functions necessary to complete the sexual cycle (Glass and Lee, 1992; Ferreira *et al.*, 1996 and 1998). Similar observations are made with mutations in *mat A-3* (Glass and Lee, 1992; Ferreira *et al.*, 1996 and 1998). MAT A-3 has sequence similarity to SMR2, another post-fertilization requiring

gene in *P. anserina* (Debuchy *et al.*, 1993). MAT A-3 has an HMG domain, characterized as a DNA binding domain in other proteins (Ferreira *et al.*, 1996; Grosschedl *et al.*, 1994).

The *mat a* idiomorph has two published ORFs. The MAT a-1 382 amino acid polypeptide confers mating-type vegetative incompatibility and mating-type identity (Staben and Yanofsky, 1990). Mutants of *mat a-1* are similar to *mat A-1* mutants; isolates are sterile and show vegetative compatibility (Griffiths and DeLange, 1978). One exception is a^{m33}, which has a missense mutation (Arg 258 → Ser) and is fertile, but is heterokaryon compatible with *mat A* strains (Staben and Yanofsky, 1990). A region of MAT a-1, as well as many other MAT a-1 equivalents in other fungi, has commonality with the HMG domain, a DNA binding motif associated with transcriptional regulation in a number of species (Staben and Yanofsky, 1990; general reference Shiu and Glass, 2000). The HMG domain in MAT a-1 has been shown to bind DNA *in vitro*, at the 5'-CTTTG-3' HMG consensus sequence (Philley and Staben, 1994). Deletion analysis has shown that this is region necessary and sufficient for mating function.

A second open reading frame in *mat a*, called *mat a-2* and encoding a putative 79 amino acid polypeptide, has recently been identified based on transcriptional RT-PCR analysis and homology in *S. macrospora* (Pöggeler and Kück, 2000). This gene has an unknown function and does not appear to play a crucial role in mating or sexual processes, and is therefore postulated to encode a non-essential protein or a pseudogene.

A comparison between the mating-type locus in *P. anserina* (Arnaise *et al.*, 1993; Glass and Nelson, 1994; Shiu and Glass, 2000) and *S. macrospora* (Pöggeler *et al.*, 1997) demonstrate conservation of mating-type identity, but not post-fertilization functions. Mating-type gene functions have been confirmed through transformation analyses involving the introduction of *N. crassa* mating-type idiomorphs into either *P. anserina* or *S. macrospora*. These transformations

restore initiation of the sexual cycle, but there is no formation of viable ascospores. The function of mating-type associated vegetative incompatibility is not conserved (Arnaise *et al.*, 1993; Pöggeler *et al.*, 1997). This indicates an evolutionary conservation of the mating-type idiomorphs in filamentous fungi for mating reactions, but not for vegetative incompatibility or post-fertilization functions as these species are evolutionarily diverse.

The only published suppressor of mating-type vegetative incompatibility is *tol* (Newmeyer, 1970; section **1.4.3.4.1**). Other potential suppressors have been identified that either partially – *vib-1* (Q. Xiang and L. Glass, unpublished results) and MH27 (this study) – or completely block vegetative incompatibility in *N. crassa* (*mak-2* – A. Pandey and L. Glass, unpublished results). *vib-1* mutations were isolated independent of mating type influences, in looking for suppressors of *het-c* vegetative incompatibility. *vib-1*, vegetative incompatibility blocked, has a number of other pleiotropic phenotypes, including partial suppression of mating-type incompatibility (Q. Xiang and L. Glass, unpublished results). *mak-2*, the *S. cerevisiae FUS3/KSS1* MAP kinase homologue in *N. crassa*, displays a constitutive conidiation and a hyphal fusion defect. *mak-2 mat a* spheroplasts transformed with *mat A-1* display a compatible phenotype and transformants maintain the ability to mate as both *mat A* and *mat a* (A. Pandey and L. Glass, unpublished results).

1.4.3.4.1 *tol*

Mating-type associated vegetative incompatibility is mediated by *tol* (*tol*erant), as demonstrated in a recessive mutation that suppresses normal mating-type vegetative incompatibility response (**Figure 7**, Newmeyer, 1970). *tol* is not linked to the mating-type locus,

and maps to linkage group IV. Mutants of *tol* are fertile, show wild type phenotype and growth rate. The only apparent phenotype of *tol* is the aforementioned suppression of mating-type associated vegetative incompatibility; it does not appear to affect other het loci (Leslie and Yamashiro, 1997). Attempts at isolating other suppressors of mating-type associated vegetative incompatibility only produced additional *tol* alleles (Vellani *et al.*, 1994). This supports the hypothesis that *tol* is a specific suppressor of the vegetative incompatibility associated with the mating-type locus.

Figure 7. The role of TOL in mating-type associated vegetative incompatibility
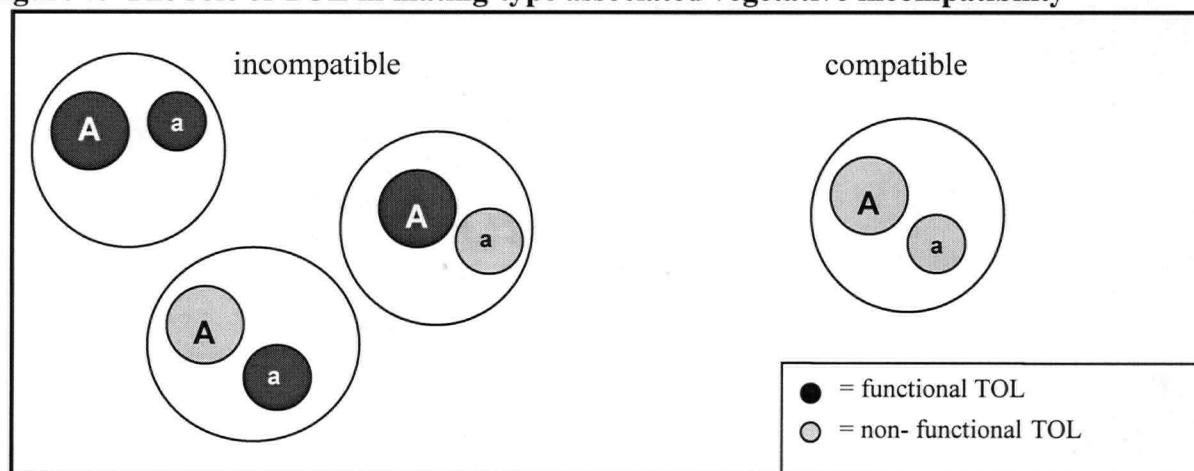
Figure 7. The influence of *tol* on mating-type vegetative incompatibility. If the forcing *mat A* and/or *mat a* strains have nuclei with a functional *tol* allele (designated by the dark circles), the resulting phenotype is a vegetative incompatibility response in the heterokaryon (large open circle). This results from functional TOL in a common cytoplasm. If *tol* is non-functional in both nuclei of the forcing strains (non-functional TOL is designated by a light grey circle), there is no functional TOL present in the cytoplasm and the heterokaryon can grow normally.

Sequences that hybridize to *tol* under a high stringency condition are present in *N. tetrasperma* and *N. sitophila*; these species do not display mating-type associated vegetative incompatibility (Shiu, 2000). Since *tol* exists in these species, it may indicate that *tol* has alternative function (Jacobson, 1992; Shiu and Glass, 1999; Shiu, 2000). *N. tetrasperma* is

pseudohomothallic filamentous fungus, which requires both mating-types for the sexual cycle, but is self-fertile; normally, *N. tetrasperma* produces four N+N (dikaryotic) ascospores, each containing one nucleus of each mating-type. tol^C introgressed into *N. tetrasperma* disrupted the normal homothallic lifestyle of *N. tetrasperma* and resulted in the segregation of *mat A* + *mat a* heterokaryotic progeny into *mat A* or *mat a* only colonies. The progeny of the introgressed strain also exhibited vegetative incompatibility. Introgression of tol^T (from *N. tetrasperma*) into *N. crassa* resulted in vegetative compatibility, similar to a *tol- N. crassa* mutant (Jacobson, 1992). Recent sequence data shows that tol^T is truncated at amino acid position 384 relative to tol^C, which correlates with the introgression analyses and the lack of function of tol^T in vegetative incompatibility (Shiu, 2000).

Recent cloning and characterization of *tol* has shown that the *tol* gene encodes a putative 1044 amino acid polypeptide, with no identified similarities to predicted protein sequences available in public databases such as GenBank (website in **Figure 10**; Shiu and Glass, 1999; modified by I. Kaneko, personal communications). The changes in sequence are a change from T to C at position 3032 and the removal of a C at position 3156 in the original *tol* sequence, which results in a frame shift and the change in polypeptide length of 1011 to 1044 aa. TOL has a leucine rich repeat (LRR) and an amphipathic α-helical (coiled-coil) domain (amino acid positions 804-823 and 176-211 respectively), both thought to be involved in protein-protein interactions (**Figure 8** – Shiu and Glass, 1999). Deletion constructs using various restriction enzyme sites and end filling were made to create a frame shift mutation of TOL, demonstrating the correct ORF. By deletion analysis, it was noted that amino acid regions 1-97 and 652-790 were non-essential for TOL function. Amino acid regions 99-521 and 802-1011 were essential for TOL function, including the amphipathic α-helix (amino acids 176-211) and the leucine rich

repeat (amino acids 804-826; Shiu and Glass, 1999; Shiu 2000). Molecular characterization of the original *tol* mutation (N83) identified a transversion of G→T at position 2666 (near the carboxy-terminus), creating a premature stop codon. This may indicate that the strain has a partial loss-of-function due to the remaining protein product, although this claim has not been investigated or substantiated. RIP (repeat induced point mutation) analysis was used to generate a number of new *tol* mutant strains. RIP *tol* mutants are phenotypically normal, except they no longer confer mating-type associated vegetative incompatibility (Shiu and Glass, 1999; Shiu, 2000). This phenotype is identical to the previous mutants obtained by escape from mating-type *mat A/mat a* incompatible heterokaryons or partial diploids (Newmeyer, 1970; Vellani *et al.*, 1994; Leslie and Yamashiro, 1997).

Figure 8. TOL
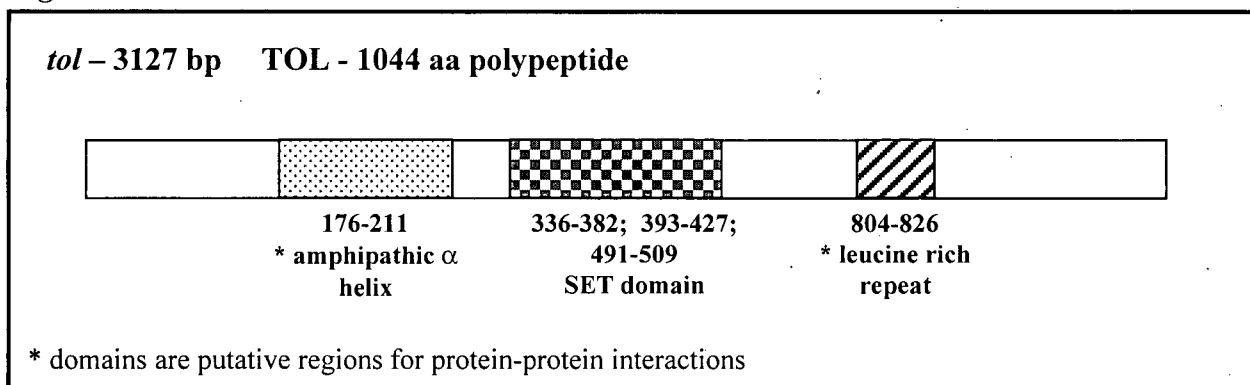
Figure 8. The putative TOL polypeptide from *N. crassa* (modified from Shiu and Glass, 1999; modified by I. Kaneko, personal communications). SET domain according to Smith *et al.* (2000b) and Saupe *et al.* (2000).

A putative SET (HET-SIX, HET-E, TOL) domain containing three regions of conservation spanning amino acids 336-509 (**Figures 5** and **8**; **Chapter 2**, this study; Smith *et al.*, 2000b; Saupe *et al.*, 2000) is common to HET-6 and TOL in *N. crassa* and HET-E in *P.*

anserina. The SET domain may confer a novel function since all three polypeptides are involved in vegetative incompatibility.

mRNA expression patterns, examined via RT-PCR, showed that *tol* is not regulated by MAT A-1 or MAT a-1 during the vegetative growth phase (Shiu and Glass, 1999). *tol* transcripts were still detected in partial diploids (*A/a tol* – D25 – **Figure 9**) despite the presence of both mating-type idiomorphs in the same organism, as well as in a mating-type deletion strain. *tol* does not appear to have a distinct role in the sexual cycle of *N. crassa*, as *tol* transcripts are not present in sexually related tissues (cross – **Figure 9**). This correlates *tol* suppression in the sexual cycle with the co-existence of *mat A* and *mat a* nuclei in the same cytoplasm for sexual reproduction.

Figure 9. RT-PCR expression analysis of *tol*
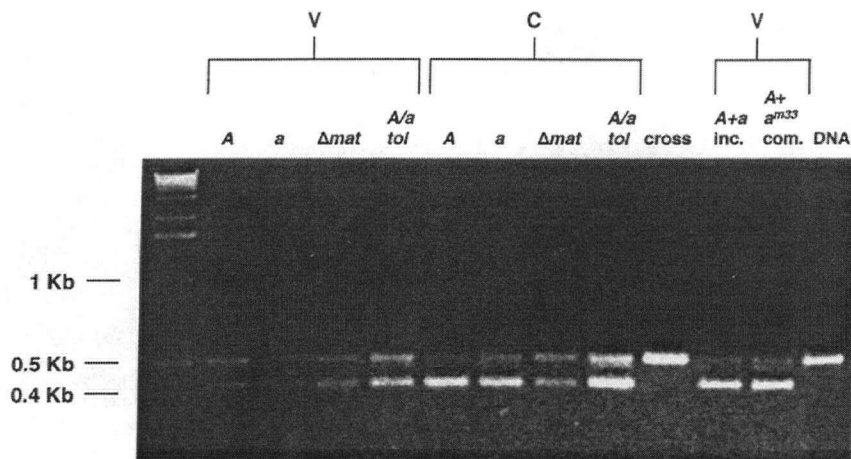
Figure 9. *tol* expression via RT-PCR in *mat A*, *mat a*, *Δmat* (deletion strain), *mat A/a tol* (partial diploid) and cross (sexual tissue) in *N. crassa*. Expression was also examined in an incompatible (*mat A + mat a*) and a compatible (*mat A + a^{m33}*) heterokaryon. Primers were tol 13 and tol 11, spanning the intron region to distinguish the size of the genomic *tol* fragment (504 bp) from the cDNA (410 bp). Strains used were: *mat A* = 74-OR23-1VA, *mat a* = OR8-1a; *Δmat* = RLM44-02; *A/a tol* = D25; *A + a* = I.20.26 + I.1.51; *A + a^{m33}* = I.20.26 + R2.11. Cultures were grown on either Vogel's (V; Vogel, 1964) or crossing media (C =Westergaard's; Westergaard and Mitchell, 1947). Modified from Shiu and Glass, 1999.

1.4.5 Mechanism of vegetative incompatibility

The current hypothesis for the mechanism of vegetative incompatibility is based on physical interaction between protein products from alternative *het* alleles, resulting in a complex that may trigger a pathway leading to incompatibility directly via a "poison" complex that results in cell death (reviewed in Glass *et al.*, 2000). This theory is supported by experimental evidence using tagged HET-C protein products (section **1.4.3.1**).

Yeast two-hybrid with MAT A-1 and MAT a-1 has established a strong *in vivo* interaction between these gene products (80 fold increase over background - Badgett and Staben, 1999). Mutations in the HMG domain of MAT a-1 that alter mating-type associated vegetative incompatibility ability but not female fertility (Staben and Yanofsky, 1990; Philley and Staben; 1994), eliminate the association with MAT A-1 (Badgett and Staben, 1999). The interaction between MAT a-1 and MAT A-1 is of interest because it may indicate that complex formation is a common trigger of vegetative incompatibility, as demonstrated with HET-C (G. Iyer, S. Sarkar, J. Wu and L. Glass, unpublished results). *het-s*, an allelic *het* locus in *P. anserina* that functions as a prion, also interacts and forms protein aggregates (Coustou *et al.*, 1997; Saupe, personal communications). This may point to a common mechanism of protein-protein interaction and recognition which triggers or causes the downstream effects of slow growth rate and cell death associated with vegetative incompatibility (Glass *et al.*, 2000). Since *tol* has putative protein-protein interaction domains, such as the leucine rich repeat and an amphipathic α-helical domain, this may indicate that TOL may work in a similar fashion.

2. COMPARATIVE ANALYSES OF TOL AND TOL-LIKE SEQUENCES

2.1 INTRODUCTION

Genetic analyses have identified approximately 10,000 genes in *N. crassa*, many of which have subsequently been cloned and characterized, resulting in more than one thousand mapped genes in the current *Neurospora crassa* compendium (Perkins *et al.*, 2001). In 1997, an EST database was established, based on mRNA expression from perithecial/sexual, conidial, mycelial samples (Nelson *et al.*, 1997). This was expanded with putative circadian rhythm regulated genes collected from evening and morning samples, in the current form of 13,000 cDNA clones (Zhu *et al.*, 2001; **Figure 10**). Genome sequencing began at a German institute (MIPS – **M**unich **I**nformation Center for **P**rotein **S**equences; **Figure 10**), sequencing linkage groups II and V. In 2000, the Whitehead Institute formed an independent sequencing project on the basis of shotgun sequencing. The entire *N. crassa* genome sequence was completed in the spring of 2001 and is available from the Whitehead Institute sequencing project in the form of contig sequences (**Figure 10**). At present, the Whitehead sequencing project consists of 1705 contigs containing 38,244,162 base pairs. The drawback of the project headed by the Whitehead Institute is that the genome has not yet been annotated (first version to be released in September, 2001 and due to be completed near the end of 2001) and so current locations of unmapped genes in the genome remain unknown.

Many public databases have been established over the last ten years (examples listed in **Figure 10**). Ones such as National Center for Biotechnology Information GenBank (NCBI; **Figure 10**) serve as a general depository for published and unpublished sequences alike, providing a wealth of sequence information in the form of DNA and protein sequences. Genes and proteins of known function help to establish common motifs in sequences, such as

localization targeting signals, DNA binding domains, RNA binding motifs and regions involved in protein-protein interactions. BLAST algorithms serve as the basis for most sequence searches of a subject sequence against the general databases. Websites that use BLAST (Altschul *et al.*, 1990), TBLASTN and BLASTP algorithms (Altschul *et al.*, 1997) are NCBI GenBank, MIPS and the Whitehead Institute. Once similar sequences are identified, alignments programs such as CLUSTALW (Thompson *et al.*, 1994) and Multalin, enable identification of similar regions that may be important to the function of the protein of interest. A popular website for access to many of the analysis tools is ExPasy, accessing many algorithms for DNA and protein manipulations, protein motifs and sequence alignments programs, all employed to determine the similarity between two or more sequences and determine putative function for genes and proteins. A general reference of all the websites used in this study can be found in **Figure 10**.

Figure 10. Websites used in this study

NCBI (GenBank) general sequence depository
http://www.ncbi.nlm.nih.gov/BLAST/

EST database for *N. crassa* (New Mexico)
http://biology.unm.edu/biology/ngp/home.html

Oklahoma site (Aspergillus; ESTs for *N. crassa* morning and evening collected samples)
http://www.genome.ou.edu/fungal.html

MIPS *N. crassa* sequencing of LGs II and V
http://www.mips.biochem.mpg.de/proj/neurospora/

The Whitehead Institute *N. crassa* genome project
http://www-genome.wi.mit.edu/annotation/fungi/neurospora/

Fungal Genetics Stock Center (FGSC)
http://www.uk.embnet.org/research/fgsc/index.html

CLUSTALW alignment program
http://www2.ebi.ac.uk/clustalw/

Multalin alignment program
http://www.toulouse.inra.fr/multalin.html

Molecular Toolkit for DNA and protein sequence manipulation
http://arbl.cvmbs.colostate.edu/molkit/

ExPasy for DNA and protein sequence manipulation
http://www.expasy.ch/

DNA sequencing facilities at UC Berkeley
http://idrive.berkeley.edu/dnaseq/web

The Neurospora homepage
http://www.neurospora.com

PSORT II
http://psort.nibb.ac.jp/form2.html

Blocks
http://www.blocks.fhcrc.org/blocks_search.html

Pfam
http://www.sanger.ac.uk/Software/Pfam/search.shtml

GCG
http://socrates.berkeley.edu:7029/gcg-bin/seqweb.cgi (user name required)

Boxshade Server
http://www.ch.embnet.org/software/BOX_form.html

DNA translation into protein
http://arbl.cvmbs.colostate.edu/molkit/

Figure 10. Websites used in this study, focusing on databases (general, *N. crassa* and Aspergillus), DNA and protein sequence manipulation programs and general *N. crassa* reference sites.

As described in Shiu (2000) and the general introduction, TOL has an amphipathic ∝-helical domain (amino acid positions 176-211) and a leucine rich repeat (amino acid positions 804-823). Alignment programs also permitted Smith *et al.* (2000b) to identify a region that shows similarity with TOL, called the SET domain for HET-SIX (*N. crassa*), HET-E (*P. anserina*) and TOL (*N. crassa*) – amino acid positions 337-353, 398-433 and 491-500 in TOL (as defined by Smith *et al.*, 2000b). The range of amino acids was modified slightly to 336-382, 395-426 and 491-509 (Saupe *et al.*, 2000). It was only after this project was started and the genome was completed, that other TOL-like proteins in *N. crassa* were discovered with similarity in the SET domain.

2.2 APPROACH AND PROCEDURES: COMPARING TOL AND TOL-LIKE SEQUENCES

Figure 11 – Outline for determining TOL-like sequences and regions of similarity

Use TOL sequence in BLAST search of Whitehead Genome sequence and EST databases

↓

Identify TOL-like sequences

↓

Align protein sequences with TOL to determine possible regions of importance, using CLUSTALW and protein motif search programs

2.2.1 Selection of sequences for analysis: putative TOL-like sequences

With the emergence of the *N. crassa* genome sequence, it was discovered that numerous TOL-like sequences existed elsewhere in the *N. crassa* genome. This chapter compares TOL with newly identified TOL-like proteins by examining regions of similarity.

The BLAST, TBLASTN and BLASTP searches were completed with the amino acid sequences of TOL (modified from the published sequence by I. Kaneko, personal communication; section **1.4.3.4.1**). The protein sequence was chosen for alignments due to the inherent conservation of amino acid sequence because of selection and function. Searches were completed using the MIPS database sequence of linkage groups II and V (website in **Figure 10**). Updated searches were subsequently performed with the Whitehead Institute genome sequence (website in **Figure 10**). A search was also made with the EST database (website in **Figure 10**) to determine if there are a number of expressed TOL-like sequences, as well as a comparison with the Aspergillus genome (website in **Figure 10**), to determine if TOL-like sequences exist in other filamentous fungi. TOL-like sequences were identified based on E values less than or equal to -1.

2.2.2 Aligning TOL and TOL-like sequences

A published region, named the SET domain (Smith *et al.*, 2000b; Saupe *et al.*, 2000), shows similarity between TOL and HET-6 in *N. crassa* and HET-E in *P. anserina*, which are known vegetative incompatibility loci. In order to determine overall regions of similarity, individual alignments were established between TOL and the TOL-like proteins. Genomic contigs were identified using BLAST of the Whitehead Institute genome sequence. Contig DNA

sequences were retrieved, with corresponding supercontig identification numbers, translated into protein sequences (website in **Figure 10**) and used in alignments (using CLUSTALW – website in **Figure 10**) with TOL and the other TOL-like sequences. Alignments were also completed including HET-6 and HET-E. These searches permitted overall comparisons of the TOL protein sequence to the TOL-like sequences to determine if there were perceptible motifs that could confer function or identify regions within sequences with known function.

In examining the SET domain, an alignment comparing TOL and all of the TOL-like sequences (including HET-6^{OR}, HET-6^{PA} and HET-E) was completed, spanning the entire region identified by Smith *et al.* (2000b) and expanded by Saupe *et al.* (2000), including sequence beyond that previously published. Alignments were then completed with TOL and all TOL-like proteins (including HET-6^{OR}, HET-6^{PA} and HET-E) for each of the three specific regions of the SET domain (**Figure 5**). The regions were narrowed (in size and number of contig reference sequences) to identify the sequences with the most conservation. In total, each SET domain region had two alignments – 1) TOL with all contig reference sequences and HET-6^{OR}, HET-6^{PA} and HET-E and 2) TOL with the narrowed set of conserved contigs and HET-6^{OR}, HET-6^{PA} and HET-E only if they maintained a conservation in sequence. Once narrowed, the conserved regions from TOL were used in protein motif scans to determine regions of known function.

BLAST searches were also completed with TOL and the MIPS database and the EST databases (websites in **Figure 10**).

2.3 RESULTS

2.3.1 Identification of TOL-like sequences

The Whitehead TBLASTN search with TOL yielded 38 contigs, including one matching the TOL sequence exactly (**Figures 12** and **13**). Five contigs 1.75 (supercontig 9), 1.1121 (supercontig 172), 1.1342 (supercontig 232), 1.301 (supercontig 34) and 1.1539 (supercontig 295) had multiple regions with similarity to the SET domain in TOL (**Figure 13**), rendering a total of 42 TOL-like protein sequences in the *N. crassa* genome (**Figure 13**). Three contigs, contig 1.765 (supercontig 107), contig 1.75 (supercontig 9) and contig 1.1342 (supercontig250) contained a sequence with low identity (42%, 45% and 35% respectively – data not shown) outside of the SET domain region. Because of the relatively low identity, this region outside the SET domain was not examined. In general, a region approximately spanning amino acids 256-650 in TOL maintains similarity with the TOL-like contigs from the *N. crassa* genome, although the SET domain ranges from 336-509 amino acids according to Smith *et al*. (2000) and Saupe *et al*. (2000) (**Figure 13**).

Figure 12. **The Whitehead Institute** *N. crassa* **genomic contigs identified in a TBLASTN with TOL**

```
Query= (1044 letters)

                                                                E
Sequences producing significant alignments:                    Value

Neurospora crassa sequence contig 1.119  (supercontig 15)      0.0
Neurospora crassa sequence contig 1.765  (supercontig 107)     e-140
Neurospora crassa sequence contig 1.1086 (supercontig 165)     2e-93
Neurospora crassa sequence contig 1.75   (supercontig 9)       3e-93
Neurospora crassa sequence contig 1.859  (supercontig 119)     5e-51
Neurospora crassa sequence contig 1.1121 (supercontig 172)     1e-42
Neurospora crassa sequence contig 1.882  (supercontig 126)     2e-42
Neurospora crassa sequence contig 1.939  (supercontig 136)     2e-41
Neurospora crassa sequence contig 1.1342 (supercontig 232)     1e-40
Neurospora crassa sequence contig 1.1414 (supercontig 250)     3e-40
Neurospora crassa sequence contig 1.926  (supercontig 133)     4e-40
Neurospora crassa sequence contig 1.301  (supercontig 34)      2e-38
Neurospora crassa sequence contig 1.1099 (supercontig 168)     3e-38
Neurospora crassa sequence contig 1.786  (supercontig 107)     2e-37
Neurospora crassa sequence contig 1.1539 (supercontig 295)     4e-36
Neurospora crassa sequence contig 1.339  (supercontig 41)      2e-35
Neurospora crassa sequence contig 1.790  (supercontig 107)     1e-34
Neurospora crassa sequence contig 1.614  (supercontig 83)      7e-32
Neurospora crassa sequence contig 1.480  (supercontig 65)      5e-30
Neurospora crassa sequence contig 1.440  (supercontig 58)      5e-30
Neurospora crassa sequence contig 1.840  (supercontig 115)     2e-28
Neurospora crassa sequence contig 1.1388 (supercontig 244)     7e-28
Neurospora crassa sequence contig 1.1375 (supercontig 239)     4e-26
Neurospora crassa sequence contig 1.588  (supercontig 77)      1e-25
Neurospora crassa sequence contig 1.946  (supercontig 136)     1e-25
Neurospora crassa sequence contig 1.409  (supercontig 52)      1e-24
Neurospora crassa sequence contig 1.322  (supercontig 38)      5e-22
Neurospora crassa sequence contig 1.1526 (supercontig 290)     1e-18
Neurospora crassa sequence contig 1.337  (supercontig 41)      8e-16
Neurospora crassa sequence contig 1.390  (supercontig 50)      1e-14
Neurospora crassa sequence contig 1.751  (supercontig 104)     8e-14
Neurospora crassa sequence contig 1.807  (supercontig 109)     3e-12
Neurospora crassa sequence contig 1.802  (supercontig 108)     1e-11
Neurospora crassa sequence contig 1.73   (supercontig 9)       3e-11
Neurospora crassa sequence contig 1.777  (supercontig 107)     6e-10
Neurospora crassa sequence contig 1.782  (supercontig 107)     1e-09
Neurospora crassa sequence contig 1.467  (supercontig 62)      5e-09
Neurospora crassa sequence contig 1.218  (supercontig 25)      7e-09
```

Figure 12. Thirty-eight contigs were identified in the TBLASTN (Altschul *et al.*, 1997) search with TOL and the *N. crassa* genomic database (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/). TOL is identified as contig 1.119 (supercontig 15). Included with the TBLASTN results are the contig and supercontigs numbers and the E values (the probability of the sequence being random), which represents the similarity of sequence in reference to TOL.

Figure 13. Regions of similarity between TOL and *N. crassa* genomic contigs

```
                                                    Alignment with TOL
                                                    (amino acid position)

         Contig 1.119   (supercontig 15)                   TOL
         Contig 1.765   (supercontig 107)          86-842 ; 938-1044
         Contig 1.1086  (supercontig 165)               273-837
         Contig 1.75    (supercontig 9)       276-837 ; 291-658 ; 979-1044
         Contig 1.859   (supercontig 119)               278-857
         Contig 1.1121  (supercontig 172)          279-658 ; 279-733
         Contig 1.882   (supercontig 126)               227-658
         Contig 1.939   (supercontig 136)               288-658
         Contig 1.1342  (supercontig 232)      279-722 ; 288-519 ; 554-657
         Contig 1.1414  (supercontig 250)               278-657
         Contig 1.926   (supercontig 133)               291-658
         Contig 1.301   (supercontig 34)            267-661 ; 279-708
         Contig 1.1099  (supercontig 168)               253-706
         Contig 1.786   (supercontig 107)               290-658
         Contig 1.1539  (supercontig 295)           284-666 ; 291-657
         Contig 1.339   (supercontig 41)                286-731
         Contig 1.790   (supercontig 107)               342-658
         Contig 1.614   (supercontig 83)                259-657
         Contig 1.480   (supercontig 65)                300-658
         Contig 1.440   (supercontig 58)                287-795
         Contig 1.840   (supercontig 115)               343-607
         Contig 1.1388  (supercontig 244)               314-656
         Contig 1.1375  (supercontig 239)               292-658
         Contig 1.588   (supercontig 77)                294-656
         Contig 1.946   (supercontig 136)               297-662
         Contig 1.409   (supercontig 52)                274-542
         Contig 1.322   (supercontig 38)                313-692
         Contig 1.1526  (supercontig 290)               292-657
         Contig 1.337   (supercontig 41)                291-526
         Contig 1.390   (supercontig 50)                290-623
         Contig 1.751   (supercontig 104)               342-656
         Contig 1.807   (supercontig 109)               246-684
         Contig 1.802   (supercontig 108)               343-660
         Contig 1.73    (supercontig 9)                 291-543
         Contig 1.777   (supercontig 107)               343-605
         Contig 1.782   (supercontig 107)               291-655
         Contig 1.467   (supercontig 62)                290-519
         Contig 1.218   (supercontig 25)                290-607
```

Figure 13. *N. crassa* genomic contigs from **Figure 12**, indicating the relative amino acid position in TOL (1011 amino acid polypeptide - Shiu and Glass, 1999).

Eleven TOL-like sequences were identified in the MIPS BLAST searches (data not shown), all with a smallest sum of probability values < E -15 located on linkage groups II or V. Since only a portion of the identified sequences encoding TOL-like proteins were found on

linkage groups II and V, it is likely that the *tol*-like genes are located randomly throughout the genome.

The *N. crassa* cDNA BLAST database searches (website in **Figure 10**) with TOL yielded no significant hits. There were also no significant hits (E < -1) with a TOL BLAST search of the Aspergillus database (website in **Figure 10**). This corresponds with the study of *tol*, where mRNA expression levels were very low and only detected using RT-PCR (Shiu and Glass, 1999; Shiu, 2000).

2.3.2 Alignment of regions of similarity between TOL and TOL-like sequences

As described in section **1.4.3.2**, the SET domain in TOL lies in three regions: amino acid positions 336-382; 395-433 and 491-509 (combined from Smith *et al.*, 2000b and Saupe *et al.*, 2000). The entire set of TOL-like protein sequences was aligned with TOL, HET-6^{OR}, HET-6^{PA} and HET-E. The alignment is viewed in **Figure 14**, with the SET regions highlighted in the TOL sequence. As mentioned in the figure legends, "w" indicates a second sequence within a contig (see **Figure 13** for contigs with multiple sequences). Manual adjustments were made in the HET-6^{OR}, HET-6^{PA} and HET-E sequences because the similarity determined by Smith *et al.*, (2000b) and Saupe *et al.*, (2000) was lost with the large volume of sequences used in the alignment. Despite the adjustments, there remain alignment differences due to the inclusion of other TOL-like sequences. Alterations were made to reduce the differences as much as possible without making changes in the rest of the data set. The three regions in the SET domain were identified among the TOL-like protein sequences, although there was degeneracy in the amount of conservation. Region three was barely identifiable, with only one semi-conservative

substitution, but more careful analysis determined a higher level of conservation (section

2.3.2.3).

Figure 14. A complete alignment of the SET domain in TOL and TOL-like sequences

⟵ Region one of the SET domain ⟶   ⟵ Region

TOL	336	LA........VSETPFMSLSHCWGK-DG--VPTQLLKGN.......VDRFTKEGIRYTELRKTFRDAIEVTQRLN----IVPYIWIDSLCYIQD	407
1.765		LG-----------FVEKSYMTLSHCWGK-NG--VPTRLMEEN------YARFLN-GIQLGQLPKTFRDAINVTRKL----KIPFIWDSLCIIQD	
1.75		MD-----------FQKQRYMTLSHCWG--DS--VPARLLLNN------YASRLK-GFALDELPRTFQDALLLTQRL----NVQFLWIDSLCIIQD	
1.1086		LD-----------LSRTDYMTLSHCWG--DG--VPLRLTHDN------YNRFLA-GIEFSEIPKTFRDAIDATRRL----SVQYLWIDSLCIVQD	
1.339		LP-----------VGHSTTVVALSYCWG--GQ--QSLTLKTEN------LEAMKL-GLMISSLPQTLQDAIAVTRAL----DQQYLWDALCIIQN	
1.588		LAK----------DLKQPQPYTALSYCWG--SG--TQAMLNDTN------LADFLK-LLPESRLSKTHTRAIDLTRLL----GIRYLWIDALCIIQK	
1.859		-------------ENTSGRYATLSHRWAS-QG--MLK-LEVNN------FEKFQQEIPWTS-LPKTYQDAIEATRRL----GIQYIWIDSLCIIQD	
1.409		-------------KGDSFRYIALSHCWGK-HE--HFF-TTALN------RRDHMRGIHVDN-LPKTFQHAVKTTRGL----GIRYLWIDSMCIIQG	
1.1099		TLK----------RQEGCINYITLSYRWTA-ET--ALTSLRTYN------KALYHDPIPET-LPQIYKDAAAVARAL----HIPYVWIDSLCIIQD	
1.614		--R----------IKDTNVKYITLSYRWTV-ET--PKTNLKTHN------KDKYHQSIPTEN-WPQIYKDAVALSRTL----GIRYVWIDSLCIIQD	
1.1539		-------------FRVDQVRYNTLSYQWTA-DT--SKTSLRKHN------RAAYCHAIPTEA-WPKVYKDVVIVCRAL----GVRYVWIDSLCIVQD	
1.882		TS-----------ENVSIQYLTLSHSWAI-TSTDSALKLTSRN------LELLQIHIPIEG-LPRTFQDALEITRRL----GYRYLWIDSLCIMQD	
1.1539w		DEG----------LAQKKAHYLTLSHSWSI-STKPHNLKLETEN-----MTQLQDRIPLDDRLSKTFRDAMKITQQL----GCRYIWIDSLCIIQS	
1.301w		AS-----------ASSPDLQYMTVSHSWHA-TG--VTEKLTTEN------ESQWLQGIPIEK-LSPNFRDAMTITRRL----GCRYLWIDSLCIIQD	
1.1342w		TS-----------PSPKPPTVMTLSHCWGL-TP--PLLTTTTT------LDSHLSQIPWSS-LPILFRDAILLVREL----GQRYLWIDSLCILQD	
1.786		KLVRSASDQIQKGGSALQYITLSHCWGP--PEKRPITTTRAN-----LSVRTE-RISFAELPETFQDAVVLTRKL----GQRYLWIDSLCIIQD	
1.840		-------------YLTLSHCWGP--PEKRPATTTRAS------LVANKN-RILFSSLPKTFQDAVEITRRL----GQRYLWIDSLCIIQD	
1.75w		RLVTSTADDLFFDGKPPKVVTLSHCWGP--PSKQPLTTTKAT------LRQRMM-RIHFEKLPQTFQDAVKICRSL----RQRFLWIDSLCIIQD	
1.790		-------------PYLTLSHCWGP--PEKRPATTTKAN------LTQRME-RIPFIELPRTFRDAIEITRKL----GHRYLWIDSLCIIQD	
1.1342		GE-----------RVLYICLSYCWGR----RPFLRTLFDN------IESHKR-CINPDDLPPTMHDAIEYTRQL----GVHYIWIDSLCIIQD	
1.1414		ST-----------LGRYTCLSHCWGG----AQPLTTTTHN------IAQHLE-RIEWNSISKTFQDAIKFTAHL----GISYIWIDSLCIIQD	
1.480		TKTP---------RAPYAALSHCWGP--PDKQPLKTKLDT------VQDHIHSGIPFDRLPPTFADAVWTREL----GIRYLWIDSLCIIQD	
1.1388		GK-----------RGYTALSYCWGSSLKAERPYLTTTKT------LEKHLS-ELPWDLLPKTLQDSIRLTRAI----GVRYLWIDSLCIIQD	
1.1121		PTT----------PYMTLTHRWGF----ADEEMLNKR------TYPSLIGGIPISSMPKLFQEVITVALHL----DVNYIWIDALCIFQD	
1.1121w		PAG----------PYATVTHRWGQ----VDEDFVLNKQ------KYPQLMKGIPLDSLPRLFQHTISVARSL----GVQYLWTDSLCIFQD	
1.939		KSH----------AYVAFSHCWGK----SEALKLLQDDLDKGTRG--NIEELRSLIOVQSLPTSYREAISVSLAL----GFRHIWIDSLCIIQN	
1.926		AHY----------PYVALSHCWGT----VSRLKLQKS------NMADLLKHIRITELPTTYREGISVTLAL----GFSYIWIDSLCIIQD	
1.322		DVDMS--------ALKFASLSYVWGVTPQKLTLLIENEY------ELQRPNSLVGKVSKTISDAIHLTEQL----GLQYLWVDALCIIQN	
1.337		QHRLG--------DIPYVTLSYVWGTT-QQAMLKRENLL------RLQLPGSLQGATPQTITDAMMFTSYM----GYRYLWDTLCIIQN	
1.390		GSTSRGGQSL---TPPEYVTLSYVWGQGPPEGPVKIKHP------SGSRQQLSLPASLPLTISDTIQVVQRL----GYRYLWIDRYCIPQD	
1.777		-------------YVTLSYVWGQ--AGEAGSHGP------VLREDAVTLPDDLPLVISDTVEVVKRL----GYRYLWIDRYCIPQN	
1.218		NATNQVAN-----YVTLSYLWGTAEATGESVVQSTN------CDGGSGLALPTQIPLVISDAIRVVKQL----GYRYLWVDRYCIPQE	
1.467		KPR----------YVTLSYVWGT--SGATSMTTP------TLPEGDQLPKIVNDAMIVTLKL----GYKYLWVDRYCIPQN	
1.807		DKA----------YLTLSYVWGTP--SGSTSGIVE------DNDSVSTLPTRLPRVVEDSIRVVKEL----GFRYLWVDRYCIPQT	
1.73		-------------DSYIALSYVWGNQPLVVNK------GSRVLPTHAPKVIEDAMVVKKL----GGRYLWVDQYCIDQH	
1.782		-------------YRFFALSYVWGKQPALKSEYALVRRGTRET------ASASGKRQLPSVVPAVIEDAMIMVTSL----GGRYLWVDQFCIDQD	
1.802		-------------YVALSYVWGTKPDDGNEGDMDIS------DRTVDWPLFSKLPPLIEDTTAVTKL----GFKYLWIDRYCIPQG	
1.751		-------------PYVALSYVWGRVKVFKTCEKDIDKLR------LRGGVEEKKFLFPRSIRDAMTLAQRI----GYCYIWIDSICIIQD	
HET-6OR	53	AP-----------ISPPPSYIALSYVWGDSTRTHEISVANEVN------DGR-AFIPLRLTSSLDTCLRHLRELHYRRQLEPLPLWIDQICINQD	128
HET-6PA	53	VP-----------ISQAPSYIALSYVWGDSTRTHEISVVNEVN------DGRGAFVTLRLTTSLDTCLRHLRTLQRQWQIAPLPLWIDQLCINQE	129
1.301		SG-----------TYITLSHCWGD--PGLMSTKLTAHN------LEEYMNEGISIEDLPRTFRDAVELTKDL----GVSYLWIDSLCIVQA	
1.1526		VDLQ---------QVEYVALSYVWGR-DPFFTLQSPNLRT------LCEKDAFRDPAIKLPQTILDAIEVTKFL----GYRYLWIDSLCIEQD	
1.440		ASES---------AKSVPYAALSYCWGR----NQKLKLEMSS------LDRFRTTGIGMEELHRTIADAVMAVLSL----GLRHMWDALCIIQD	
1.1375		VP-----------YVVLSYSWGDPTTMPAEEWARIKGASTKTVDGVPAPERTNPFPRSILPETMRDAIAITQELN-----FKYIWIDNVCIPKG	
HET-E	18	LP-----------SGKIPPYAILSHTWGP-----DEEEVSYKD------LKDGRAVSKLGYNKIRFCADQAWRD------GRKFFWDTCCIDKS	84

:   . .:: *   :   . .* * *** :

two of the SET domain ⟶

```
TOL       SKEDWDDESVKMQYVARNSVKNLAAGASPNSHGGLFNPRHPLSTVPWSIEVPLSDNDDKDYNI-----------------471
1.765     -------SYEDWMIQESAKMQQVYQNSFLNLAAGASLNSKGGLFYRRHPLSVVPWSVRLGANGYLTAK-----------
1.75      ----------SPDDWVAESAKMHFVYQNTHLNLAAAVSPNSSGGFFPRHPLSFVPWLMRLDTYS----------------
1.1086    ----------SREDWLHESAKMHHIYQNSYLNLMAAASSNHGGLYSSKYPFLSIPFLVPLGDPRNPKIAFYR---------
1.339     ----------SPSDWEVESSKMASIYRGAYLTIAAATASDVTQGFLSGNYRETNAWYKEPYHEEWINQQDCSTI--------
1.588     ----------NQKDWEAESKRMSSVYGNAELTIIAGRSPDSGLGPLTNNYLTTKKHPKPVPMRAGPN-----PH--------
1.859     S---------ADGKDWIAESAKMTSVYENSHINFAAADAVDSTQGCFFDRDPSSIRPLRLRIAVEHKSE------------
1.409     ----------PGGDFDKQSAKMEDVFSSAYCVIAASSARGQGDGFLN--PRRKCDHVKVVNESAKGA--------------
1.1099    ----------STQDWNEQASLMDSIYTHARLNLSGMFA-DRAVGLRVERDPLAVSPCIISRRLPTDCDAD-----------
1.614     ----------DGKDWNTQASLMHRIYTHGYLNLAHACA-EFSPGLEVTRDPISVSPCVLSRTRTASSS-------------
1.1539    ----------DLEDWKVQSSMMQDVYSHGHLNLANVLG-KHADGLEVSRYPTATSPCILSRTLEDGST-------------
1.882     ----------SESDWHREALSMSSIYGNSSCNIAALGV-DGADTCFRQRNPLQVIPCQITQQDGTS---------------
1.1539w   GI--------KAGDWKREGLTMSDVYGHSRCNITALGT-GGNDGCYTPRNPLQTSPCHIKVCDDKPA--------------
1.301w    ----------SREDWEREVLTMTNVYGHSTCNIAMVGQ-AGDEGCFRKRNPLVAFPCRLHEADGKESRLY-----------
1.1342w   ----------SESDWLTESANMSSIYAHATLNIAATAQ-RDASVSLFRQRFHQQGFRELRVSAAGRS--------------
1.786     ----------DENDWAQEASTMVDVYTQSYCTLAALSSGDSTAGLCKK----DQTNPDMLIDISVDDGHNEPFNGRLVS---
1.840     ----------DEADWAHEASLMAKVTHSYCTLAALSSSDSSQGLN-------MVRPDFLLNHEIRGNQNDQY----------
1.75w     ----------DEGDWAKEAALMASVYSYSYFTLAALDSKDSSGGCK-------MVADIQSSYSNRFVDIDFYG--LEE-----
1.790     ----------DEQDWAYEAALMAKIYSHAFCMLSALSSNDSSEGLRLEPLDEDSSYMDLMTTSHASPAGGSGIIKSDMD---
1.1342    ----------SVLDWREEAGKMASIYQRSYLVLSATSSADAYGMHVS---LPSDYQTFALTVNTRTN---EICDSNIA----
1.1414    ----------SPQDWVEESAKMCSIYENGHLTIAASASNGSGGLPIR---DSSGFVRLSGTTSSPGS--KPFD---------
1.480     ----------SASDWETESQKMGFIYEGATLVLGAAHARDSSEGLFKG---LQRPDPKKVVKVYTDMAPYAT----------
1.1388    ----------SNDDWEKEAARMGTLYAQARLVIAATTGADAEAGLFP-----PQTYIYRPSETS-----------------
1.1121    K---------DDQTDWQNEAGOMQKVYSNSFCNISAADATSCSDTLFSQPRNLKEELLPQTIYLSIPSSSSSSSSSFPSSLSPTFTTSTASQTSSRL
1.1121w   K---------DDKTDWAREASLMEKVYSYSYCNISAADAENCSWSLFNARD-PQSINPETLTLDIVSDDIGG----------
1.939     ----------SRDDWAKESAMMQDVYANSSLNLCASAAADSSKASFQRRDLGVYVPLKIEPRWTGILDGDHKKGKPGPL----
1.926     ----------DEKDWAEEAAMMKDVFEHCFINLSATAAPDSLQSNFTHRDATSILLEGET-----PNRRGEKHGPS------
1.946     ----------DQEDLREELNHMPEVYQYATLTICAASSTSIYDGFLYRRGYWTHGCRPMNLRVAADCSES-----------
1.322     S---------DDDKVMQLGNIANVYAHSLFTIMAAAGDDSNFGLPGIRTPRDTVQEVVPVIGPSQNN--------------
1.337     D---------DADKMSQLQIMGDIYKNANFTIVAAAGENSGSGLAGILTPRRAIQRKVQVRPAGPQE--------------
1.390     D---------LPAKQIQENMGHIYSRSVLTIIAAAGEGPEYGLPGVSERWRAEQLTVQVGAAAQ-----------------
1.777     D---------SPVKHLQIRKMDRIYSCSVLTIIAAAGDGPEYGLPGVSSRHRTEQLSVQV--------------------
1.218     D---------GAAKHRQILNMGRIYSNSSLTIIAAAGDEPEYGLPGVSSRSRLAQVGVQIN-------------------
1.467     D---------PYAKHIQIQNMGSIYSLSLITIGAAGDAEYGLPGVSSAPRVPQLWTDIPYNGG------------------
1.807     D---------SRAKHTQIQLMGTIYARSAITIVAAAGKDAEYGLPGVDARERNPQLRIDLERVDGNL--------------
1.73      D---------DQDKHAQIKNMDRIYEGSHATIVAFSGVDSSSGLPGVSSTPRIPQLRFTS---------------------
1.782     D---------PDDKHAQIGNMAAIFEGAYATIVAFSATDSASGLPGVRNLARKPYPSVTH---------------------
1.802     D---------QEAKHTQIQNMDSIYGSASLTIIAAAGNDPSHGLPGVS-LPRYGRGRPDKDPPFSVK--------------
1.751     DDTD------TYQNLDKRHQIGQMGKIYQYADLTIVAASGEDANAGLPGVEAGTREARQKEVKIS---------------
HET-6OR   D---------NEEKSFQVRLMRDIYSSAHQVVWLGPAVDDSNRVMDALAEVGQEFLDKIG----------------180
HET-6PA   D---------DAEKSSQVLLMKNIYSSAHQVVWLGPAADGSDKLMDAFVEIGQGFLDKLG----------------181
1.301     DPSKDMESHAKMVREDWERESASMCTIFAGSVTLGAMASPDCQGGLFAPVKQRLISFDTMPDGSHS--------------
1.1526    S---------TEDTVTQIGSMHVIYAQASLTIVAAVGDNARHSILHINPNQLSNKYHTIGGLRFCLD-------------
1.440     S---------DEDKAREIGNMGAIYAAATVTLAVSGAASVDDGFLCKRSLGNLGLWSFEIPTQTARTSLQIPVTDAGAGKR-
1.1375    ----------TDWDTEASLMHEVYGNAAFTLLASSSTKATEPMLYDRLAWLQEPKPCKLRNHFLYN--------------
HET-E     N---------STELQEAINSMFRWYRDAAKCYVYLTDVSTDKRDADGDPSWKWAFQKCKWFT-------------137
```

(cont.)

Region three of the
SET domain

```
TOL       ----------KNKTF----LTSEYRSEKESDLIER[TRGWVLQEQLLARRTEL]----  509
1.765     ------------HISEYCHERYAKLLIN--------TRGWVLQEQLLARRTLVFGRK
1.75      -----------TLALHYELRGPSILY--------RRGWVLQEQLPRRTLIFGRQ
1.1086    -----------YMNARREKLDNLTLF--------SRGWVMQERVLARRNLIFG-K
1.339     -------LGARLKYGVFGHDDERHRVSALPWN--------LRGWTLQEQFLSRRLLTYHAY
1.588     -------IGEGPDLGTVHLTMQRAKS-VGPVD--------TRGWCFQEALLSNRALVFGED
1.859     ----------YYYAATSHRFAYNVAPAPLY--------RRGWVYQERTLAPRIIHCCAK
1.409     ----------VYVSRFRDDFKEHVLNSPLS--------ERGWVLQERALARRTIYFTEW
1.1099    ---EEWYEHFACFEGYGANSFVDLAPLF--------RRAWCHQERYLARRTVYFG-E
1.614     ------EYWACFVNVTPSSDLARAPLY--------SRGWCYQERFLATRTVHFS-Q
1.1539    ------EHWGCMEDGSWREHIQDAPLY--------TRGWCYQERFLSARTVQFG-Q
1.882     ----------IYAGPKLELDT--------PLL--------RRGWVMQERLLSPRVLYAGAA
1.1539w   -------IYAVNSLLENTFTENVEQSPLF--------KRGWVFQESVLSSRAVYFGAP
1.301w    ----A-FIEGSNRKFSRGPEEIVGSFPLL--------KRGWVVQEALLSPRTIFFGDH
1.342w    ------------------------------------AAGLGLRAPMDTVEIGSLPLS
1.786     ----KPNVDSE--YDGWWISEPIGTECRPTNPLR--------YRAWALQEKELSTRVVHFGFR
1.840     ----RVQAWKGT--RSGWSGEYYT--SRDRSPLR--------YRAWTLQERELSTRTIHFGHQ
1.75w     ----GGTKFPRQ--RLRVFEVPSLEDGSIEVGHLN--------SRAWTLQERELSRRVISFASQ
1.790     ----ESFIWFRISIHLEAWGIIYGGTSYHDTISPLC--------TRAWTLQESRLSRRVIYFAKH
1.1342    --DQANFPMTNTTETFYVRRAYDHVNRGTSSYAGRIPMLPTNKRGWIFQERFLSSRVLHFGPH
1.1414    ----IIGYFELGGIEARDHPSPTNHRTPLYS--------RGWVHQEMLLSPRVVHFAKK
1.480     ----CFYTVPEWPYVMELEPWNNSR-LST--------RAWVMQEFTLSRRFVWSEKN
1.1388    ----WSLYFRKYQKERGSLQHNLPLFT--------RGWAFQEWVLARRIVFWAQD
1.1121    LSSILNHHSIPKPPYIAMNRPKIKFEPFTLINLSLWTHQVSHALVN--------TRAWVLQERVLAPRILHFGRR
1.1121w   ------MIP----VQFYVVNPAFWFQQVSNALVN--------TRAWVLQERVLAPRVLHFGKR
1.939     ------PSYLHRIKPKLVIYGMHEMEINNSPLN--------LRAWVVQERILSRRHLFMTNH
1.926     ------RFWFARS----WVDLRQEDIVHSPLH--------SRAWFQEANLARRRIDLARS
1.946     ------FILFVYMHLAFGVFDLRKDPIF--------KRAWTYQEYVLSPHILSYRTH
1.322     ------PGMSLITALSP-THQSHEHPTRKTIWA--------SRGWTLQERALSRRAIVIMKS
1.337     ------MPLWLISTVMPRTGSSNRSYTHSLPWQ--------TRGWTLQEISLSRRVFIFTGE
1.390     ------GKGISLALYER--PRVAIT-DSKWH--------TRGWTFQEGLLSRRRLVFTDQ
1.777     ------TEEISLVFYGK--PRALVA-ASKWN--------TRGWTYQEGLLSKRRLVFTDE
1.218     ------DKISLVLYEP----PTDHIT-NSKWN--------TRGWTYQEGLLSKRRLVFTDL
1.467     ------NNHLQLIYFNP--PDTEIR-TAKWH--------SRGWTYQEGLLSKRRLVFSDR
1.807     ------RLPPSLMVYRL--PQDDIR-SSLWS--------TRGWTYQEALLSKRRLVFTDN
1.73      ------SKVTLLGFTP--FLTRQTLE-ASAWM--------KRGWTFQEALLSRRLLIFTQE
1.782     ------LKHPISALRPSAISSAKLVT-ASVWM--------TRGWTYQEAILSRRLLIVTDY
1.802     ------IGSRTFVTVGTG-DMPRQLR-ESKWN--------NRAWTMQEAFLSCRCLVFFDD
1.751     ------EDLGIVNRSVA--WPDALL-KTTWE--------ERGWTYQERILSLRYMYFVDD
HET-6OR   -DHTEEEHWLSVDRLIKEKIEQPDAVTFLREAYKVIYMLNREHSFTRWVERTWFKRLWTIQEFCLCADTIF----  250
HET-6PA   -DHTTEEYLLSVRRLIEKNIEQPGVVAFLRQSYKVVYALKRDGYFARWEKRPWFTRVWTIQEFCLCSDTVF----  251
1.301     ----YLYAEEHHNHDEKYLLES--------RAWVYQETLVSPRSLHFGLR
1.1526    ------RPEFKEVVASSTWA--------TRGWTLQEITCSDRFIFFTPE
1.440     EKGKFSQEANVEPKTTALLLHKSPNGGQIRWIVQNASPLYTRAWTYQERLLSARILDVGVI
1.1375    ------TQKPLHEVRLDPPAS--------QRAWTLQEERLSPRILYWTGQ
HET-E     ----------LFTRGWTLQELIAPTSVEF----  153
```

Figure 14. The entire SET domain, spanning aa positions 336 – 509 of TOL. Amino acid residues are also numbered for HET-6 (OR and PA) and HET-E. The highlighted regions in TOL are the three regions identified by SMITH *et al.* (2000b), modified by SAUPE *et al.* (2000). Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Alignment output was modified for HET-6 (OR and PA), to show alignment in the third region in the SET domain.

Region one of the SET domain (amino acids 336-382) has 2% identity (1/47) and 11% positivity (5/47), where identity = identical amino acid residues / total number of residues x 100% and positivity = identical and conserved amino acid residues / total number of residues x 100%. The terms identity and positivity are defined in the results outputs from BLAST using the NCBI GenBank database search program (website in **Figure 10**). Region two of the SET domain spans amino acids 395-433 and has 10% identity (4/39) and 21% positivity (8/39). The third region of the SET domain (amino acids 491-509), as mentioned previously, is only identified by one semi-conservative amino acid substitution, giving a positivity value of 5% (1/19).

A BLASTP search with the entire SET region of TOL was performed, with an increased E value from the standard 10 to 10,000. The results (data not shown) did not identify proteins with identity or positivity values over 75%, except those identified as TOL or TOL-related proteins. Because the individual regions in the SET domain yielded more interesting proteins in their BLASTP searches, focus will be made regarding these results.

2.3.2.1 Region one of the SET domain

As mentioned previously, there are three identified regions within the SET domain (Smith *et al.*, 2000b; Saupe *et al.*, 2000). An alignment of TOL with all the TOL-like sequences and HET-6 (OR and PA) and HET-E was compiled (**Figure 15a**). All 42 TOL-like sequences from the genomic contigs were included in the alignment. There are some changes between this alignment and that of **Figure 14**, which contained the entire SET region. This is due to the variations with additional data and volume of sequence.

Figure 15a. Region one of the SET domain

```
TOL        343  FMSLSHCWG 351
1.1342w         YMTLSHCWG
1.1086          YMTLSHCWG
1.765           YMTLSHCWG
1.75            YMTLSHCWG
HET-E       26  YAILSHTWG   34
1.480           YAALSHCWG
1.301           YITLSHCWG
1.786           YITLSHCWG
1.840           YLTLSHCWG
1.75w           YVTLSHCWG
1.790           YLTLSHCWG
1.1342          YICLSYCWG
1.1414          YTCLSHCWG
1.1388          YTALSYCWG
1.588           YTALSYCWG
1.339           YVALSYCWG
1.859           YATLSHRWA
1.409           YIALSHCWG
1.882           YLTLSHSWA
1.1539w         YLTLSHSWS
1.301w          YMTVSHSWH
1.1539          YNTLSYQWT
1.614           YITLSYRWT
1.1099          YITLSYRWT
1.1121          YMTLTHRWG
1.1121w         YATVTHRWG
1.939           YVAFSHCWG
1.926           YVALSHCWG
1.1375          YVVLSYSWG
1.946           WCALSYVWG
HET-6OR     60  YIALSYVWG   68
1.802           YVALSYVWG
1.751           YVALSYVWG
HET-6PA     60  YIALSYVWG   68
1.73            YIALSYVWG
1.782           FFALSYVWG
1.390           YVTLSYVWG
1.777           YVTLSYVWG
1.218           YVTLSYLWG
1.467           YVTLSYVWG
1.807           YLTLSYVWG
1.322           FASLSYVWG
1.337           YVTLSYVWG
1.1526          YVALSYVWG
1.440           YAALSYCWG

                :   . :: *
```

Figure 15a. The first region in the SET domain, spanning aa positions 343 – 351 of TOL. Amino acid residues are also numbered for HET-6 (OR and PA) and HET-E. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX _form.html).

To determine if there was more conserved amino acid residues from the one identical, three conservative substitutions and one semi-conservative substitution in the general alignment, an alignment of a narrowed region with fewer TOL-like sequences was completed (**Figure 15b**). Choice for exclusion was based on deviance from a general consensus at a particular location, until a region with a higher conservation was established with TOL and 30 of the TOL-like sequences. In this alignment, there are two identical amino acid residues and four conservative substitutions. Similarity of the region was maintained with HET-6 (OR and PA) and HET-E. The final region is much smaller than that previously published, spanning amino acids 343-351 in TOL versus 337-353 (Smith *et al.*, 2000b) or 336-382 (Saupe *et al.*, 2000). The order of the contigs in **Figure 15b** differed from **Figure 14**, due to the removal of TOL-like contig sequences and the use of a narrowed region of TOL. **Figure 15b** also represents a different alignment pattern versus that seen in **Figure 14**, as the new alignment based the consensus on identity of the TOL-like contig sequences instead of the overall TOL sequence.

Figure 15b. Region one of the SET domain: reduced number of genomic contig sequences

```
TOL        343 FMSLSHCWG 351
1.1342w        YMTLSHCWG
1.75           YMTLSHCWG
1.1086         YMTLSHCWG
1.765          YMTLSHCWG
1.409          YIALSHCWG
1.1342         YICLSYCWG
1.1414         YTCLSHCWG
HET-E       26 YAILSHTWG 34
1.480          YAALSHCWG
1.301          YITLSHCWG
1.786          YITLSHCWG
1.840          YLTLSHCWG
1.790          YLTLSHCWG
1.75w          YVTLSHCWG
1.1121         YMTLTHRWG
1.1121w        YATVTHRWG
1.926          YVALSHCWG
1.339          YVALSYCWG
1.946          WCALSYVWG
HET-6OR     60 YIALSYVWG 68
1.73           YIALSYVWG
HET-6PA     60 YIALSYVWG 68
1.802          YVALSYVWG
1.751          YVALSYVWG
1.390          YVTLSYVWG
1.777          YVTLSYVWG
1.218          YVTLSYLWG
1.467          YVTLSYVWG
1.807          YLTLSYVWG
1.1526         YVALSYVWG
1.322          FASLSYVWG
1.337          YVTLSYVWG
1.440          YAALSYCWG

               :  :::  **
```

Figure 15b. The first region in the SET domain, spanning aa positions 343 – 351 of TOL, aligned with a narrowed number of 30 TOL-like genomic contigs. Amino acid reference numbers for HET-6 (OR and PA) and HET-E are also indicated. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conserved substitutions; "." = indicates semi-conserved substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

The first region in the SET domain appears to span nine amino acids (343-351 in TOL), with two conservative residues and four semi-conservative amino acid substitutions. A BLASTP search, with an increased E value parameter (E 10,000 versus the standard 10), was completed with the narrowed TOL sequence against the NCBI GenBank. Most hits had very high E values (data not shown), but TOL and TOL-like sequences were identified, as were UDP-N-acteylglucosamine pyrophosphorylases from humans, a putative adenylate cyclase regulatory

45

protein in *Trypanosoma brucei* and RNA-directed RNA polymerases from numerous viruses

(**Figure 16**). In general, identity values were above 65% and positivity values were above 85%

(**Figure 16**). Protein motif searches were unsuccessful because the region of interest is likely too

small for a comprehensive comparison.

Figure 16. BLASTP with region one of the SET domain

```
                                                           %      %
Sequences producing significant alignments:               Id.    Pos.

(9 aa)
gi|7493976|pir||T17430   tol protein - Neurospora crassa   100    100
gi|11359682|pir||T50979   related to tol protein            77     99

(7 aa)
gi|14424722|gb|AAH09377.1|AAH09377 UDP-N-acteylglucosamine
    pyrophosphorylase Homo sapiens                          85     85
gi|10720387|sp|Q16222|UAP1_HUMAN UDP-N-acteylglucosamine
    pyrophosphorylase                                       85     85
gi|4507759|ref|NP_003106.1| UDP-N-acteylglucosamine
    pyrophosphorylase H. sapiens                            85     85

(7 aa)
gi|9366571|emb|CAB95333.1| possible putative adenylate cyclase
    regulatory protein Trypanosoma brucei                   71     99
gi|9366566|emb|CAB95328.1| possible putative adenylate cyclase
    regulatory protein T. brucei                            71     99
gi|9366582|emb|CAB95344.1| possible putative adenylate cyclase
    regulatory protein T. brucei                            71     99

gi|4927037|gb|AAD32996.1| RNA-directed RNA polymerase [porcine
    transmissible gastroenteritis virus]                    75     75
gi|4927025|gb|AAD32990.1| RNA-directed RNA polymerase
    [canine coronavirus]                                    75     75
gi|4927027|gb|AAD32991.1| RNA-directed RNA polymerase [feline
    infectious peritonitis virus]                           75     75

(6 aa)
gi|13475828|ref|NP_107398.1| ABC transporter, periplasmic   66     99
    binding-protein [Mesorhizobium loti]
```

Figure 16. BLASTP (Altschul *et al.*, 1997) results of a search of the NCBI GenBank database with the first region of the SET domain (TOL sequence 343-351 aa). GenBank reference numbers, protein descriptions, % Id (identity = # identical amino acid residues/total number of residues x 100%) and % Pos (positivity = identical and conservative amino acid residues/ total number of residues x 100%) with TOL are indicated. Hits are grouped according to similar protein function and the aa (amino acid) span aligning with TOL is indicated above each group. Only those hits with % values = or > 75 are included in this figure.

2.3.2.2 Region two of the SET domain

As described for the first region in the SET domain, an alignment of TOL with the 42 TOL-like genomic contig sequences, HET-6 (OR and PA) and HET-E was completed (**Figure 17a**). Reasons for alteration in contig order and alignment pattern from **Figure 14** have been mentioned in section **2.3.2.1**. Another possibility for alterations in alignment consensus from the published SET domains (Smith *et al.*, 2000b; Saupe *et al.*, 2000) may be degeneracy or lack of sequence from insertion/deletion events in SET region.

Figure 17a. Region two of the SET domain

```
TOL        395 PYIWIDSLCIIQD------------SKEDWDDESVKMQYVYR 424
1.765          PFIWIDSLCIIQD------------SYEDWIQESAKMQQVYQ
1.1086         QYLWIDSLCIVQD------------SREDWLHESAKMHHIYQ
1.75           QFLWIDSLCIIQD------------SPDDWVAESAKMHFVYQ
1.859          QYIWIDSLCIIQDS----------ADGKDWIAESAKMTSVYE
1.409          RYLWIDSMCIIQG------------PGGDFDKQSAKMEDVFS
1.1099         PYVWIDSLCIIQD------------STQDWNEQASLMDSIYT
1.614          RYVWIDSLCIIQD------------DGKDWNTQASLMHRIYT
1.1539         RYVWIDSLCIVQD------------DLEDWKVQSSMMQDVYS
1.882          RYLWIDSLCIMQD------------SESDWHREALSMSSIYG
1.1539w        RYIWIDSLCIIQSGI---------KAGDDWKREGLTMSDVYG
1.301w         RYLWIDSLCIIQD------------SREDWEREVLTMTNVYG
1.1342w        RYLWIDSLCILQD------------SESDWLTESANMSSIYA
1.786          RYLWIDSLCIIQDD-----------ENDWAQEASTMVDVYT
1.840          RYLWIDSLCIIQDD-----------EADWAHEASLMAKVYT
1.75w          RFLWIDSLCIIQDD-----------EGDWAKEAALMASVYS
1.790          RYLWIDSLCIIQDD-----------EQDWAYEAALMAKIYS
1.301          SYLWIDSLCIVQADPSKDMESHAKMVREDWERESASMCTIFA
1.480          RYLWIDSLCIIQDS-----------ASDWETESQKMGFIYE
1.1388         RYLWIDSLCIIQDS-----------NDDWEKEAARMGTLYA
1.1342         HYIWIDSLCIIQDS-----------VLDWREEAGKMASIYQ
1.1414         SYIWIDSLCIIQDS-----------PQDWVEESAKMCSIYE
1.339          QYLWVDALCIIQNS-----------PSDWEVESSKMASIYR
1.588          RYLWIDALCIIQKN-----------QKDWEAESKRMSSVYG
1.1121         NYIWIDALCIFQDKD----------DQTDWQNEAGQMQKVYS
1.1121w        QYLWTDSLCIFQDKD----------DKTDWAREASLMEKVYS
1.939          RHIWIDSLCIIQN------------SRDDWAKESAMMQDVYA
1.926          SYIWIDSLCIIQD------------DEKDWAEEAAMMKDVFE
1.1375         KYIWIDNVCIPKG------------TDWDTEASLMHEVYG
1.946          RYLWIDALCIIQDD-----------QEDLREELNHMPEVYQ
het-6OR    116 -PLWIDQICINQD------------DNEEKSFQVRLMRDIYS 145
het-6PA    117 -PLWIDQLCINQE------------DDAEKSSQVLLMKNIYS 146
1.802          KYLWIDRYCIPQGDQEA--------------KHTQIQNMDSIYG
1.751          CYIWIDSICIIQDDDTDYQ------NLDKRHQIGQMGKIYQ
1.390          RYLWIDRYCIPQDDLPA-------------KQIQIENMGHIYS
1.777          RYLWIDRYCIPQNDSPV-------------KHLQIRKMDRIYS
1.218          RYLWVDRYCIPQEDGAA-------------KHRQILNMGRIYS
1.467          KYLWVDRYCIPQNDPYA-------------KHIQIQNMGSIYS
1.807          RYLWVDRYCIPQTDSRA-------------KHTQIQLMGTIYA
1.73           RYLWVDQYCIDQHDDQD-------------KHAQIKNMDRIYE
1.782          RYLWVDQFCIDQDDPDD-------------KHAQIGNMAAIFE
1.322          QYLWVDALCIIQNSDDD-------------KVMQLGNIANVYA
1.337          RYLWVDTLCIIQNDDAD-------------KMSQLQIMGDIYK
1.1526         RYIWIDSLCIEQDSTED-------------TVTQIGSMHVIYA
het-e       73 -FFWVDTCCIDKSNSTE-------------LQEAINSMFRWYR 101
1.440          RHMWVDALCIIQDSDED-------------KAREIGNMGAIYA
                 .* *   ** :                        :      :
```

Figure 17a. The second region in the SET domain, spanning aa positions 395 – 424 of TOL. Amino acid residues are also numbered for HET-6 (OR and PA) and HET-E. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

48

The general alignment in **Figure 17a** yielded four identical amino acids and four conservative amino acid substitutions over a range of 30 amino acids (TOL as reference). In contrast, an alignment was completed using an extended region of TOL and 24 TOL-like genomic contig sequences (**Figure 17b**). In this alignment, HET-6 (OR and PA) and HET-E were among the excluded sequences. This is the most significant region in common among TOL and the TOL-like sequences due to its size and level of identity. The region is still present in HET-6 (OR and PA) and HET-E, as evidenced in **Figure 17a**, though not with the same degree of conservation as other TOL-like sequences.

Figure 17b. Region two of the SET domain: extended region of TOL and 24 TOL-like sequences, excluding HET-6 and HET-E
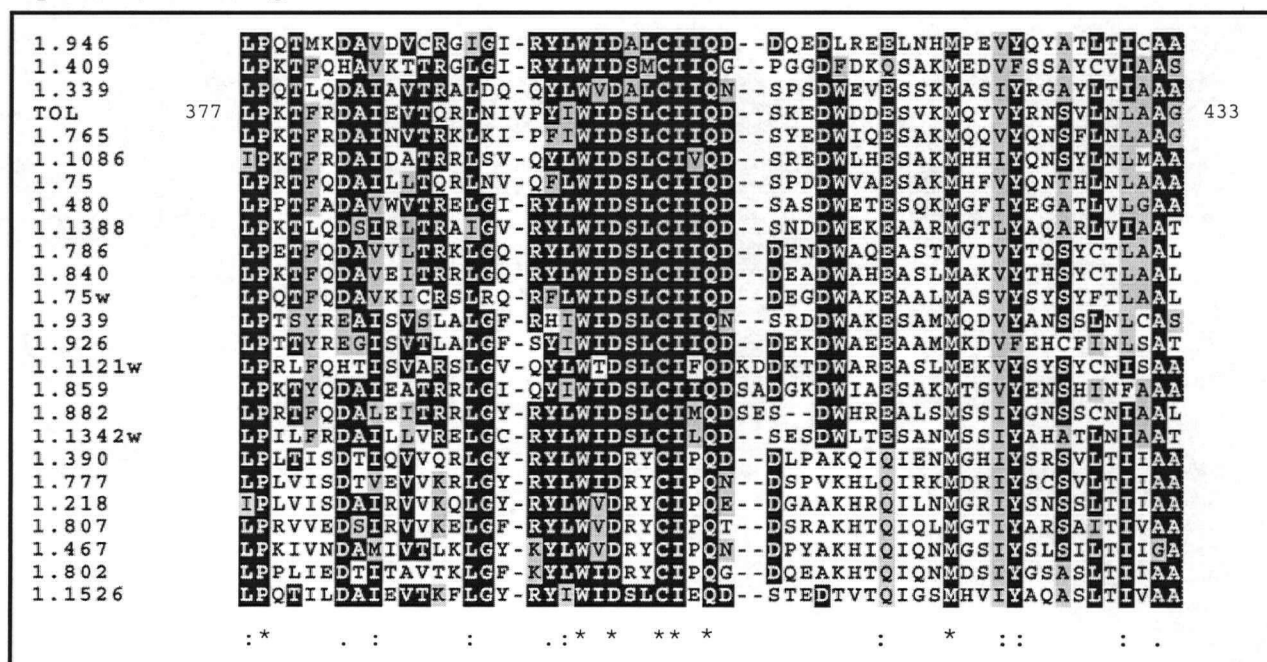
Figure 17b. The second region in the SET domain, spanning aa positions 377 – 433 of TOL, aligned with a narrowed number of contigs and excluding HET-6 (OR and PA) and HET-E. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

The extended version of the second SET domain region spans 57 amino acids (377-433) in TOL, with seven identical, eight conservative and three semi-conservative amino acid residues substitutions out of a total 57. The extended region was used in an NCBI GenBank BLASTP database search, with an increased E value of 10,000 versus the standard value of 10 (**Figure 18**). The BLASTP results included in **Figure 18** represent only the hits with an E value equal to or less than 43 (this value was chosen due to the repetition of proteins with a known function). As expected, TOL and TOL-like sequences were identified including HET-6 and a HET-6-like protein, but also included were two sequences for a protein involved in DNA repair in *S. cerevisiae*, Mms21p. The identity was 27% and the positivity 57%. This was comparable to values indicated for TOL-like proteins (which ranged from 37-66% identity and 49-83% positivity).

Figure 18. BLASTP with region two of the SET domain: extended TOL region

```
                                                            # aa sim.    %      %
Sequences producing significant alignments:                 to TOL     Id.    Pos.


gi|7493976|pir||T17430   tol protein - Neurospora crassa      57       100    100
gi|11359682|pir||T50979  related to tol protein [imported]    56        66     83
gi|11359684|pir||T51209  related to tol protein [imported]    57        50     65
gi|11359680|pir||T49475  related to tol protein [imported]    54        44     69
gi|11359681|pir||T49614  related to tol protein [imported]    56        50     64
gi|11359683|pir||T51057  related to tol protein [imported]    56        37     49


gi|6320817|ref|NP_010896.1|  involved in DNA repair; Mms21p   43        27     57
gi|585490|sp|P38632|MM21_YEAST  DNA REPAIR PROTEIN MMS21      43        27     57


gi|11595729|emb|CAC18207.1|  related to het-6 protein         42        26     56
gi|11359394|pir||T51076  het-6 protein                        51        29     48


gi|7468226|pir||F72041  efflux protein - Chlamydophila
    pneumoniae                                                21        38     66


gi|13111418|dbj|BAB32803.1|  RNA-dependent RNA polymerase
    Porcine enterovirus                                       22        50     68
gi|13111422|dbj|BAB32805.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111406|dbj|BAB32797.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111420|dbj|BAB32804.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111416|dbj|BAB32802.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111400|dbj|BAB32794.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111424|dbj|BAB32806.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111404|dbj|BAB32796.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
gi|13111410|dbj|BAB32799.1|  RNA-dependent RNA polymerase
    P. enterovirus                                            22        50     68
```
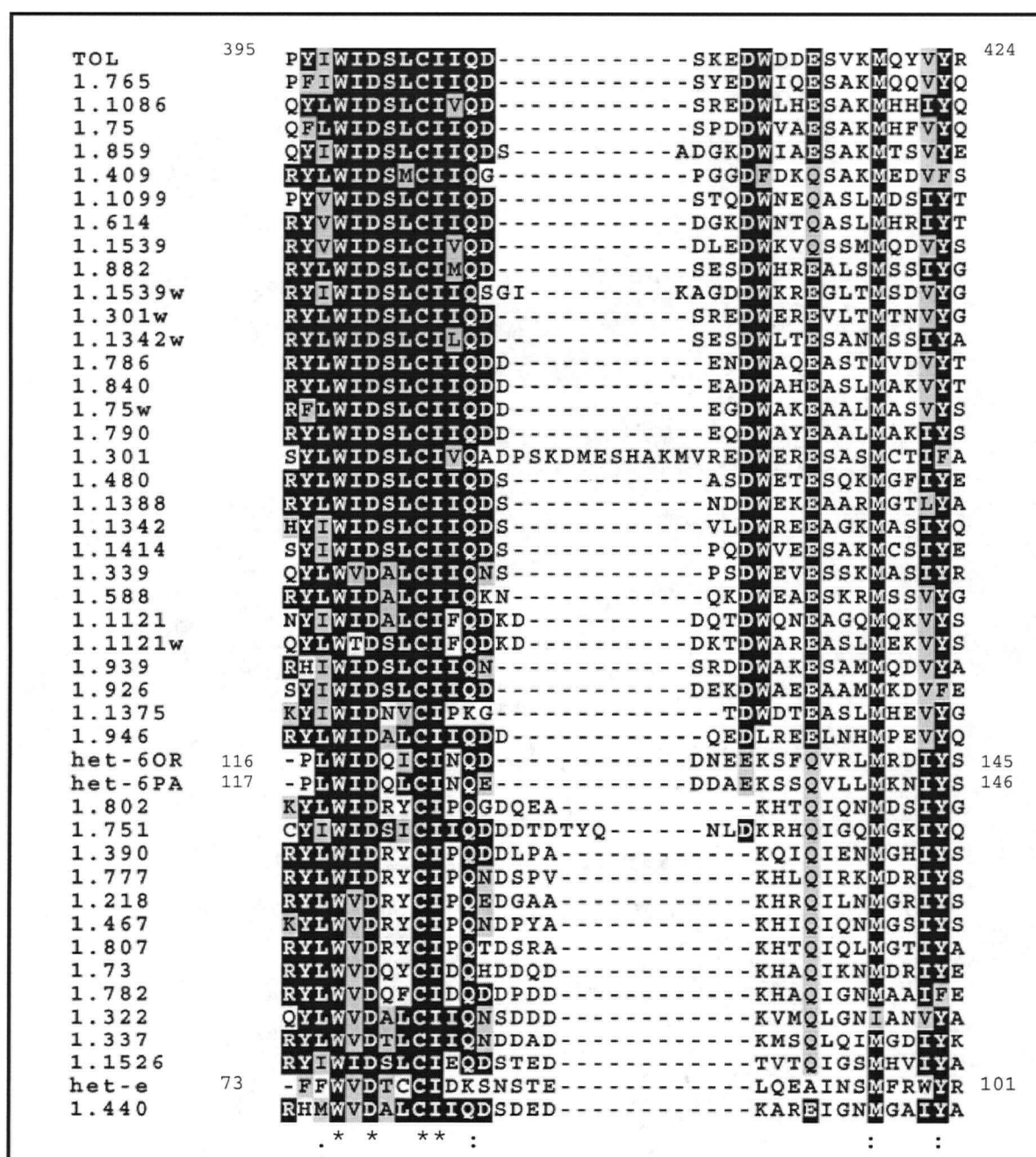
Figure 18. BLASTP (Altschul *et al.*, 1997) results of a search of the NCBI GenBank database with the extended version of the second region of the SET domain (TOL sequence 377-433 aa). GenBank reference numbers, protein descriptions, # of amino acids (aa) similar to TOL, % Id (identity = identical amino acid residues / total number of residues x 100%) and % Pos (positivity = identical + conserved amino acid residues / total number of residues x 100%) with TOL are indicated. Hits are grouped according to similar protein function. All hits described as "putative", "hypothetical", "probable" or "polypeptide" were excluded from the above figure. Only those hits with E values of 45 or smaller were examined with reference to TOL (data not shown).

Little is published on Mms21p, but the name is derived from **m**ethyl **m**ethane**s**ulfonate sensitive mutants (Prakash and Prakash, 1977). *mms21* mutants have an increased susceptibility to spontaneous mitotic recombination (Prakash and Prakash, 1977). These strains are also sensitive to UV and X-rays and display pleiotropic effects (Montelone and Koelliker, 1995). An alignment of the region in Mms21p with TOL is found in **Figure 19**. The significance of the similarity with such a wide variety of proteins is unclear. Protein motif scans did not yield any regions with known function.

Figure 19. Alignment of TOL and Mms21p

```
   TOL    378 PKTFRDAIEVTQRLNIVPYIWIDSLCIIQDSKEDWDDESVKMQ 420
   Mms21   134 PATMVNNTDTLKILKVLPYIWNDPTCVIPDLQNPADEDDLQIE 175

           *  *:  :    :.  :  * :::**** *. *:* *  ::   *::.:::
```
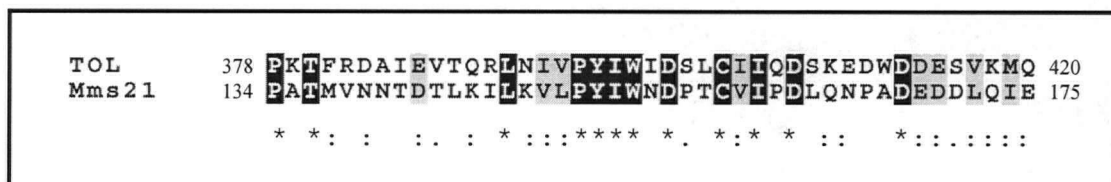
Figure 19. ClustalW sequence alignment (http://www2.ebi.ac.uk/clustalw/) of TOL (aa 378-420 of the extended second region in the SET domain) and Mms21p (GenBank access number: gi|6320817|ref|NP_010896.1|). "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

2.3.2.3 Region three of the SET domain

The third region of the SET domain identified by Smith *et al.*, (2000) and Saupe *et al.*, (2000) was aligned with TOL, HET-6 (OR and PA) and HET-E, excluding one of the two sequences from the 1.1342 contig that contained the SET domain region (**Figure 20a**). In total, 41 TOL-like sequences were aligned from the genome contigs. Alteration in alignment patterns from **Figure 14** were mentioned previously. With alignments using a shorter, more specific

region of TOL, the resulting alignments are based more on identity between the remaining

sequences.

Figure 20a. Region three of the SET domain

TOL	491 LFTR----------GWVLQEQLLARRTLI	509
1.765	LNTR----------GWVLQEQLLARRTLV	
1.75	LYRR----------GWVLQEQILPRRTLI	
1.1086	LFSR----------GWVMQERVLARRNLI	
1.339	WNLR----------GWTLQEQFLSRRLLT	
1.588	VDTR----------GWCFQEALLSNRALV	
1.1099	LFRR----------AWCHQERVLPIRTVY	
1.614	LYSR----------GWCYQERFLATRTVH	
1.1539	LYTR----------GWCYQERFLSARTVQ	
1.882	LLRR----------GWVMQERLLSPRVLY	
1.1539w	LFKR----------GWVFQESVLSSRAVY	
1.301w	LLKR----------GWVVQEALLSPRTIF	
1.859	LYRR----------GWVYQERTLAPRIIH	
1.409	LSER----------GWVLQERALARRTIY	
1.786	NPLR---------YRAWALQEKELSTRVVH	
1.840	SPLR---------YRAWTLQERELSTRTIH	
1.75w	GHLN---------SRAWTLQERELSRRVIS	
1.790	SPLC---------TRAWTLQESRLSRRVIY	
1.1342	SSYAGRIPMLPTNKRGWIFQERFLSSRVLH	
1.1414	PLYS---------RGWVHQEMLLSPRVVH	
1.301	LESR----------AWVYQETLVSPRSLH	
1.480	LSTR----------AWVMQEFTLSRRFVW	
1.1388	LFTR----------GWAFQEWVLARRIVF	
1.1121	VNTR----------AWVLQERVLAPRILH	
1.1121w	VNTR----------AWVLQERVLAPRVLH	
1.939	LNLR----------AWVVQERILSRRHLF	
1.926	LHSR----------AWVFQEANLARRRID	
1.1375	ASQR----------AWTLQEERLSPRILY	
1.946	IFKR----------AWTYQEYVLSPHILS	
HET-6OR	232 WFKR----------LWTIQEFCLCADTIF	250
HET-6PA	233 WFTR----------VWTIQEFCLCSDTVF	251
1.322	135 WASR----------GWTLQERALSRRAIV	153
HET-E	WFTR----------GWTLQELIAPTSVEF	
1.1526	WATR----------GWTLQEITCS-----	
1.337	WQTR----------GWTLQEISLS-----	
1.390	WHTR----------GWTFQEGLLSRRRLV	
1.777	WNTR----------GWTYQEGLLSKRRLV	
1.218	WNTR----------GWTYQEGLLSKRRLV	
1.467	WHSR----------GWTYQEGLLSKRRLV	
1.807	WSTR----------GWTYQEALLSKRRLV	
1.73	WMKR----------GWTFQEALLSRRLLI	
1.782	WMTR----------GWTYQEAILSRRLLI	
1.802	WNNR----------AWTMQBAFLSCRCLV	
1.751	WEER----------GWTYQERILSLRYMY	
1.440	LYTR----------AWTYQERLLSARILD	

*   **

Figure 20a. The third region in the SET domain, spanning aa positions 491 – 509 of TOL, including HET-6 (OR and PA) and HET-E, but excluding one of the two of the sequences in the genomic contig, 1.1342 that contains the SET domain region. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

As with the other SET domain regions, selective alignments were completed with TOL. When the pool of genomic contigs was reduced to determine regions of conservation, two classes

emerged. One alignment contained the TOL sequence and 22 genomic contigs (**Figure 20b**). The other alignment contained HET-6 (OR and PA), HET-E and 13 genomic contigs (**Figure 20c**). The separation between the two classes may be superficial, but since current analyses are hypothetical, no strong conclusions can be made about the difference between the two sets of sequence.

Figure 20b. Region three of the SET domain: TOL and 22 TOL-like sequences

```
1.1121           VNTRAWVLQERVLAPRILH
1.1121w          VNTRAWVLQERVLAPRVLH
1.939            LNLRAWVVQERILSRRHLF
1.926            LHSRAWVFQEANLARRRID
TOL        491   LFTRGWVLQEQLLARRTLI   509
1.765            LNTRGWVLQEQLLARRTLV
1.75             LYRRGWVLQEQILPRRTLI
1.1086           LFSRGWVMQERVLARRNLI
1.409            LSERGWVLQERALARRTIY
1.859            LYRRGWVYQERTLAPRIIH
1.786            LRYRAWALQEKELSTRVVH
1.840            LRYRAWTLQERELSTRTIH
1.1099           LFRRAWCHQERVLPIRTVY
1.614            LYSRGWCYQERFLATRTVH
1.1539           LYTRGWCYQERFLSARTVQ
1.882            LLRRGWVMQERLLSPRVLY
1.1539w          LFKRGWVFQESVLSSRAVY
1.301w           LLKRGWVVQEALLSPRTIF
1.480            LSTRAWVMQEFTLSRRFVW
1.1388           LFTRGWAFQEWVLARRIVF
1.588            VDTRGWCFQEALLSNRALV
1.440            LYTRAWTYQERLLSARILD
1.946            IFKRAWTYQEYVLSPHILS

                 :  *.*   **   *.  : :
```

Figure 20b. The third region in the SET domain, spanning aa positions 491 – 509 of TOL, narrowed to include TOL and 22 TOL-like genomic contig sequences. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/softw are/BOX_form.html).

54

Figure 20c. Region three of the SET domain: HET-6 (OR and PA), HET-E and 13 TOL-like sequences

```
1.322          WASRGWTLQE
1.337          WQTRGWTLQE
HET-E     135  WFTRGWTLQE  144
1.1526         WATRGWTLQE
1.339          WNLRGWTLQE
HET-6OR   232  WFKRLWTIQE  241
HET-6PA   233  WFTRVWTIQE  242
1.782          WMTRGWTYQE
1.390          WHTRGWTFQE
1.777          WNTRGWTYQE
1.218          WNTRGWTYQE
1.467          WHSRGWTYQE
1.807          WSTRGWTYQE
1.73           WMKRGWTFQE
1.802          WNNRAWTMQE
1.751          WEERGWTYQE

               *   *  ** **
```
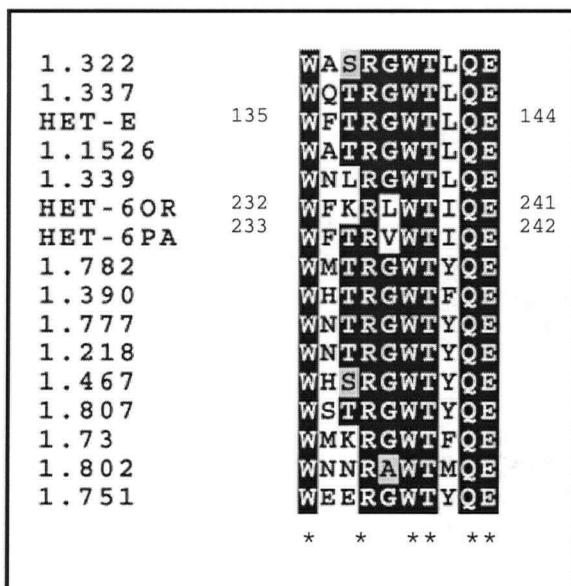
Figure 20c. The third region in the SET domain, spanning aa positions 232-241 and 233-242 in HET-6 OR and PA, respectively. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "*" = identical or conserved residues in all sequences in the alignment. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/).
Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

The third region in the SET domain is perhaps the most dramatic in comparing the alignment of TOL with the entire set of genomic contigs (excluding one of the two SET domain-containing sequences from the 1.1342 contig) with the narrowed down number of genomic contigs. In the overall alignment (**Figure 20a**), there were three identical amino acid residues (3/19 = 16% conservation, using TOL as the reference). In the TOL pool of alignments (**Figure 20b**), there are five identical, three conservative and one semi-conservative amino acid residues, spanning a 19 amino acid stretch (amino acid positions 491-509 in TOL; 9/19 = 47% similarity). In the second pool, containing HET-6 (OR and PA) and HET-E (**Figure 20c**), the number of amino acids was much smaller (ten versus 19), but there was a higher conservation, with six identical amino acid residues (6/10 = 60% identity versus 47% similarity). Between the two alignments, four of the amino acids were conserved (R, W, Q and E at positions 494, 496, 499 and 500 respectively, relative to TOL).

A BLASTP search with contig 1.1121, with an increased E value of 10,000 versus the standard value of 10, of the NCBI GenBank database, was completed with the 19 amino acid sequence of the third region in the SET domain (the class containing TOL - **Figure 21**). A number of interesting classes of proteins were identified. Since a large number of hits were made (in all, 100 hits were identified), only those of interest and having positivity values over 70% were examined. The large number of interesting proteins identified in the BLASTP search grouped as a variable region in immunoglobin kappa light chain, with an overall identity of 57% and positivity of 71%. An alignment with one of the light chain protein sequences with the sequence from the 1.1221 contig can be found in **Figure 22**. BLASTP searches with TOL yielded hits with a large variety of proteins with variable functions (data not shown). Because of the large pool of similar hits with the TOL-like sequence of contig 1.1121 and the prevalence of hits of the variable region of the immunoglobin kappa light chain (**Figure 21**), this was the sequence used for analysis.

Figure 21. BLASTP results with the third region of the SET domain: TOL-included region

```
                                                             # aa      %       %
Sequences producing significant alignments:                 ref.     Id.     Pos.

gi|11359681|pir||T49614 related to tol protein               16       68      80
gi|7493976|pir||T17430 tol protein - Neurospora crassa       16       68      80
gi|11359682|pir||T50979 related to tol protein               16       68      80
gi|11359680|pir||T49475 related to tol protein               17       52      63
gi|11359683|pir||T51057 related to tol protein               17       52      63


gi|7657092|ref|NP_055255.1| Interleukin-1 Superfamily 1
    Homo sapiens                                             13       53      76


gi|732748|emb|CAA84392.1| antibody, light chain variable
    region to HIV1 H. sapiens                                17       52      63
gi|1353828|gb|AAB01805.1| Ig light chain variable region
    H. sapiens                                               14       57      71
gi|4323894|gb|AAD16586.1| immunoglobulin kappa light chain
    variable region H. sapiens                               14       57      71
gi|470442|gb|AAA17576.1| Ig kappa chain VKIII-JK5
    H. sapiens                                               14       57      71
gi|4378250|gb|AAD19460.1| immunoglobulin kappa light chain
    variable region H. sapiens                               14       57      71
gi|10637162|emb|CAC10818.1| immunoglobulin kappa chain
    variable region H. sapiens                               14       57      71
gi|87865|pir||B25521 Ig kappa chain precursor V region
    H. sapiens                                               14       57      71
gi|306958|gb|AAA16944.1| immunoglobulin kappa chain
    H. sapiens                                               14       57      71
gi|4323928|gb|AAD16603.1| immunoglobulin kappa light chain
    variable region H. sapiens                               14       57      71
gi|722420|gb|AAA99317.1| immunoglobulin kappa chain
    H. sapiens                                               14       57      71
gi|470614|gb|AAA17700.1| Ig kappa chain VKIII-JK3
    H. sapiens                                               14       57      71
gi|4378330|gb|AAD19500.1| immunoglobulin kappa light chain
    variable region H. sapiens                               14       57      71
gi|470616|gb|AAA17701.1| Ig kappa chain VKIII-JK3
    H. sapiens                                               14       57      71
gi|1813656|gb|AAB41731.1| immunoglobulin light chain
     H. sapiens                                              14       57      71
gi|8777879|gb|AAF79136.1| immunoglobulin light chain
    variable region H. sapiens                               14       57      71
gi|1864119|gb|AAB48607.1| Ig light chain variable domain
    H. sapiens                                               17       47      64
gi|7650454|gb|AAF66049.1| immunoglobulin kappa light chain
    variable region H. sapiens                               17       47      58
```

Figure 21. BLASTP (Altschul *et al.*, 1997) results of a search of the NCBI GenBank database with the TOL-containing third region of the SET domain (genomic contig 1.1121). GenBank reference numbers, protein descriptions, number of amino acids (aa) to determine similarity, % Id (identity = identical amino acid residues / total number of residues x 100%) and % Pos (positivity = identical + conservative amino acid residues / total number of residues x 100%) with TOL are indicated. Hits are grouped according to similar protein function. Only those hits of interest and with general positivity values over 70% were included in the above figure.

Figure 22. Alignment of contig 1.1121 and an immunoglobin kappa light chain variable region

```
     1.1211          AWVLQERVLAPRIL
     Igkappa         AWYQQKRGQAPRLL

                     **    *.*    ***.*
```
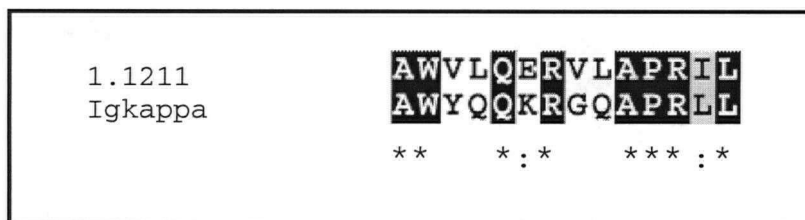
Figure 22. ClustalW sequence alignment (http://www2.ebi.ac.uk/clustalw/) of genomic contig 1.1211 sequence (the third region in the SET domain) and an immunoglobin kappa light chain variable region from *Homo sapiens* (GenBank access number: gi|4323894|gb|AAD16586.1|). "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions. Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

The BLASTP search, with an increased E value from 10 to 10,000 with the third region of the SET domain (version containing HET-6 and HET-E) differed from that of the region containing the TOL sequence (data not shown). The reason for the variation is likely due to the aforementioned smaller amino acid range and increased conservation among the amino acid residues. In this BLASTP search, the expected HET-6 and HET-E-1 sequences were retrieved, but none of the TOL or TOL-like sequences were among these identified. There was no strong conservation of the region with known proteins, as most of the hits were with putative or hypothetical proteins, or had positivity values less than 70%. Protein motif scans of both versions of the third region of the SET domain did not yield any information about putative function based on sequences of characterized proteins.

2.3.2.4 TOL and TOL-like sequences containing all three regions in the SET domain

In a final comparison of TOL and TOL-like sequences containing all three SET domain regions, a final alignment was compiled of ten sequences including TOL (**Figure 23**). Contig 1.1121w, with all three SET domain regions, was excluded from the final compilation in order to increase the overall identity and positivity of the alignment. There are variations in the overall alignment in **Figure 23** versus the individual regions of the SET domain and that found in **Figure 14** due to the modified number of contigs included in the alignment and span of the protein amino acid sequence.

Figure 23. TOL and TOL-like sequences with all three regions in the SET domain

```
TOL    343 FMSLSHCWGK-DGVPTQLLKGNYDRFTKEGIRLTEEPKTFRDAEVTQRLNEVPYIWIDSLCIIQDSKEDWDDESVKMQVVAYRNSVL 428
1.765      YMTLSHCWGK-NGVPTRLMEENYARFLN-GIQLGQLPKTFRDAINVTRKL-KIPFIWIDSLCIIQDSYEDWIQESAKMQQVYQNSFL
1.1086     YMTLSHCWG--DGVPLRLTHDNYNRFLA-GIEFSEIPKTFRDAIDATRRL-SVQYLWIDSLCIVQDSREDWLHESAKMHHIYQNSYL
1.75       YMTLSHCWG--DSVPARLLLNNYASRLK-GFALDELPRTFQDAILLTQRL-NVQFLWIDSLCIIQDSPDDWVAESAKMHFVYQNTHL
1.480      YAALSHCWGPPDKQPLKTKLDTVQDHIHSGIPFDRLPPTFADAVWVTREL-GIRYLWIDSLCIIQDSASDWETESQKMGFIYEGATL
1.786      YITLSHCWGPPEKRPITTTRANLSVRTE-RISFAELPETFQDAVVLTRKL-GQRYLWIDSLCIIQDDENDWAQEASTMVDVYTQSYC
1.840      YLTLSHCWGPPEKRPATTTRASLVANKN-RILFSSLPKTFQDAVEITRRL-GQRYLWIDSLCIIQDDEADWAHEASLMAKVYTHSYC
1.926      YVALSHCWG--TVSRLKLQKSNMADLLK-HIRITELPTTYREGISVTLAL-GFSYIWIDSLCIIQDDEKDWAEEAAMKDVFEHCFI
1.409      YIALSHCWG--KHEHFFTTALNRRDHMR-GIHVDNLPKTFQHAVKTTRGL-GIRYLWIDSMCIIQGPGGDFDKQSAKMEDVFSSAYC
1.946      WCALSYVWG--GDVPLKTTKATQFIHKE-RIPIDNLPQTMKDAVDVCRGI-GIRYLWIDALCIIQDDQEDLREELNHMPEVYQYATL
           : :**: **                .        :      : :* *       : :* **           *   :   :

TOL    429 NLAAGASPNSHGGLFN----PRHPLSTVPWSIEVPLSDDNDDKDYNIKNKTF-------LTSEYRSEKESDLIEETRGWVLQEQLLA 504
1.765      NLAAGASLNSKGGLFY----RRHPLSVVPWSVRLGANGYLTAK----------HISEYCHERYAKLILNTRGWVLQEQLLA
1.1086     NLMAAASSNSHGGLYS----SKYPFLSIPFLVPLGDPRNPKIAFY----------RYMNARREKLDNLTLFSRGWVMQERVLA
1.75       NLAAAVSPNSSGGFFF----PRHPLSFVPWLMRLDTYS---------------TLALHYELRGPSILYRGWVLQEQILP
1.480      VLGAAHARDSSEGLFKGL-QRPDPRKKVVKVYTDMAPYATCFYTVP-----E-------WPYVMELEPWNNSRLSTRAWVMQEFTLS
1.786      TLAALSSGDSTAGLCKK---DQTNPDMLIDISVDDGHNEPFNGRLVSKPNVWDSEYDGWWISEPIGTECRPTNPLRYRAWALQEKELS
1.840      TLAALSSSDSSQGLN-----MVRPDFLLNHEIRGNQNDQY-----RVQAWKGTRSGWSGEYYT---SRDRSPLRYRAWTLQERELS
1.926      NLSATAAPDSLQSNFTHR-DATSILLLEGETPNRRGEKHGPSRFWFAR--------------SWVDLRQEDIVHSPLHSRAWVFQEANLA
1.409      VIAASSARGQGDGFLNPR-RKCDHVKVVNESAKGAVYVSR--------------FRDDFKEHVLNSPLSERGWVLQERALA
1.946      TICAASSTSIYDGFLYRRGYWTHGCRPMNLRVAADCSESFILFVY--------------MHLAFGVFDLRKDPIFKRAWTYQEYVLS
           :  *         .        :: .  **  .  .:  .::   :.  :  .  . *: :. .**  . **

TOL    505 RRTLLIFGKEELHWECVTCEASESFPSSIDRERWDG-DMNDRRTIFQHQWENLTGT----DTGNKLGPADSSDMNSKRRKAWELLVQT 586
1.765      RRTLVFGRKELHWECSTCQASETFPDPKNRP-----TPKWVETISRRDWQFLLG------DKL------VDSKRQEAWNRLTRT
1.1086     RRNLIFG-KEIHWECHDSTGSESFP------------------VGTALEKRQHTVEG------QTDASRRTVWQNIVKN
1.75       RRTLIFGRQELYWECSMSKASEFFPGAGSVTWALSGTIDMQSTSVPPALSGYLGNGVF-TSSRMKLFQERLLDAERLEAWASLVAE
1.480      RRFVWSEKNSLHWVCREEMEG----------------------EPHLDTYYFDMIRS-------FGTLTRIWHFLVDD
1.786      TRVVHFGFRQLLWECCELKATAQLPWYHGVRN-----RTRRRPQGSILPLPLS--------VRMSARWLKICED
1.840      TRTIHFGHQQLLWECMEGKACTQHPWPLFDRGQ----MTLRKAVASTSEIVVYR-------PKEMRMKWQNLVEE
1.926      RRRIDLARSQMWHCQQHWACETHPSGVWSPEDYEWETEESQQYYNMCKRRDNHNASLGMSSLSGPITSHTWFEGDYRRDWLGKIHR
1.409      RRTIYFTEWQTYWECGDGIRCES------------------------------LTRMDK
1.946      PHILSYRTHHLEWKCRTHTELVGTPEFELPITKTSMTSEKIDGCEKGQQKVLMTN-------LSTYRLSSSTSWNDVVQE
           ::         * *
```

60

(Cont.)

```
TOL      587 YFSRSLTKASDRLIAISGIA  606
1.765        YFERSLTHASDRLVAISGLA
1.1086       YSRLNLTFASDRLIAIAGMA
1.75         YSGRNLTKQSDKLVAISGLA
1.480        YCRKQLTYETDRLAALHGIA
1.786        YSSRSLTKETDKLIALSGIA
1.840        YSGRSLTISTDKLIALSGIA
1.926        YSSCDITKTKDRLIAFSGIA
1.409        YVNPQFYKMSAREFWMRDAN
1.946        YTTRAASLASDKLRALSAVA
             *        .  :  : .
```

Figure 23. The compiled alignment of TOL (from aa residue 343-606) and the most similar TOL-like sequences from the Whitehead Institute genomic contigs of *N. crassa*. Alignment was completed with CLUSTALW (http://www2.ebi.ac.uk/clustalw/). The SET domain (as determined in this study) is highlighted, with regions spanning 343-351, 377-433 and 491-509 aa respectively. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions.

Comparing **Figure 14** with **Figure 23**, the two alignments spanning all three regions of the SET domain, identity and positivity values are shown to increase when TOL is aligned with fewer contig sequences. In the final overall alignment, HET-6 (OR and PA) and HET-E are omitted. For region one of the SET domain (amino acid positions 343-351 relative to TOL), the identity is 44% (4/9) and the positivity value increases from 11% to 78% (7/9), when compared with **Figure 14**. Positivity values are being compared because it combines identical amino acid residues and conservative and semi-conservative amino acid substitutions. Region two of the SET domain spans amino acid positions 377-433, with an identity value of 19% (11/57) and a positivity value of 46% versus 21% observed in **Figure 14**. The final region of the SET domain is found at amino acid positions 491-509. The identity value for the third region in the SET domain in reference to TOL is 26% (5/19) and has a positivity value of 58% (11/19). 58% contrasts with the 5% positivity value seen in **Figure 14**.

A BLASTP search with the region of TOL from amino acid 343-509 is shown in **Figure 24**. Again, the similarity with Mms21p was evident, as well as similarity to another protein involved in DNA repair, a Rad2p from *S. cerevisiae*. A recent submission to the GenBank database, a transcriptional regulator of sugar metabolism in *Clostridium acetobutylicum* showed 75% positivity with the second region in the SET domain. There is no published information about this protein and the significance of the role of a transcriptional regulator is unclear.

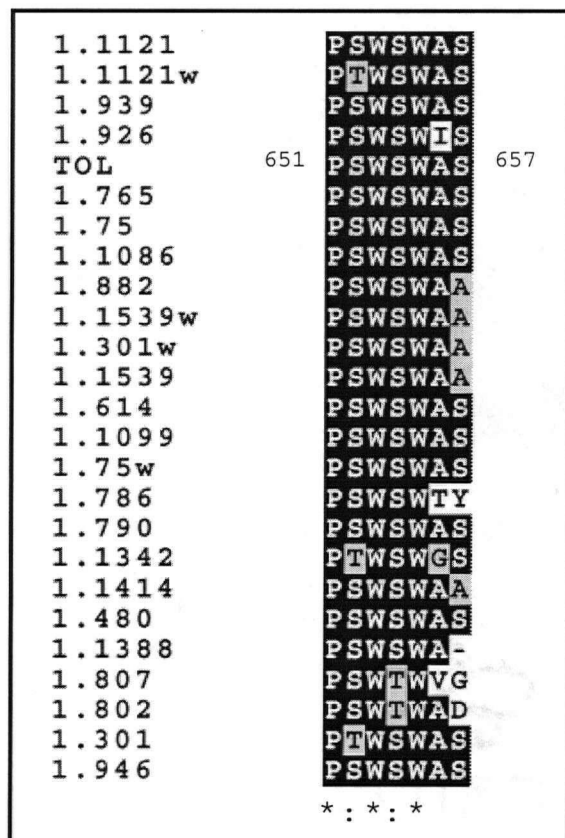**Figure 24. BLASTP results with TOL (aa 343-509)**

```
                                                              #    %    %
Sequences producing significant alignments:                 aa   Id   Pos

gi|7493976|pir||T17430   tol protein - Neurospora crassa     254   95   95
gi|11359682|pir||T50979   related to tol protein            254   46   60
gi|11359684|pir||T51209   related to tol protein      ◦     289   32   44
gi|11359680|pir||T49475   related to tol protein            271   30   46
gi|11359681|pir||T49614   related to tol protein            279   26   39
gi|11359683|pir||T51057   related to tol protein            270   25   40


gi|6320817|ref|NP_010896.1|  involved in DNA repair; Mms21p   46   28   56
    S. cerevisiae
gi|585490|sp|P38632|MM21_YEAST   DNA REPAIR PROTEIN MMS21      46   28   56
    S. cerevisiae
gi|6321697|ref|NP_011774.1|  DNA repair protein RAD2p S. cerevisiae  57   31   52


gi|11359394|pir||T51076 heterokaryon incompatibility protein het-6   88   27   46
gi|6578961|gb|AAF18153.1|  heterokaryon incompatibility protein het-6   90   26   47


gi|15024343|gb|AAK79372.1|AE007652_2   Transcriptional regulator   24   50   75
    of sugar metabolism Clostridium acetobutylicum
```

Figure 24. BLASTP (Altschul *et al.*, 1997) results of a search of the NCBI GenBank database with TOL amino acid sequence 343-509. GenBank reference numbers, protein descriptions, # of amino acids (aa) in alignment, % Id (identity = identical amino acid residues / total number of residues x 100%) and % Pos (positivity = identical + conserved amino acid residues / total number of residues x 100%) with TOL are indicated. Hits are grouped according to similar protein function. All hits described as "putative", "hypothetical", "probable" or "polypeptide" were excluded from the above figure. Only those hits of interest are included.

2.3.2.5 A putative novel domain in TOL and TOL-like sequences

In aligning TOL with 31 of the TOL-like sequences, it was discovered that a region spanning seven amino acids had a high conservation with three identical amino acid residues and two conservative and one semi-conservative substitutions (5/7 = 71% positivity; **Figure 25**). The entire set of TOL-like contigs was not included because some of them did not extend as far as this region, which is found at amino acid positions 651-657 in the TOL sequence. This region was also determined to be non-essential in deletion analyses of TOL (Shiu and Glass, 1999; Shiu, 2000).

Figure 25. A putative novel domain in TOL: an alignment between TOL and 21 TOL-like genomic contig sequences


```
1.1121          PSWSWAS
1.1121w         PTWSWAS
1.939           PSWSWAS
1.926           PSWSWIS
TOL        651  PSWSWAS  657
1.765           PSWSWAS
1.75            PSWSWAS
1.1086          PSWSWAS
1.882           PSWSWAA
1.1539w         PSWSWAA
1.301w          PSWSWAA
1.1539          PSWSWAA
1.614           PSWSWAS
1.1099          PSWSWAS
1.75w           PSWSWAS
1.786           PSWSWTY
1.790           PSWSWAS
1.1342          PTWSWGS
1.1414          PSWSWAA
1.480           PSWSWAS
1.1388          PSWSWA-
1.807           PSWTWVG
1.802           PSWTWAD
1.301           PTWSWAS
1.946           PSWSWAS

              *:*:*
```

Figure 25. A putative novel domain between TOL and 21 TOL-like genomic contig sequences, spanning aa positions 651-657 in TOL. Contigs are named according to the main contig number (supercontig numbers are omitted for simplicity). "w" denotes a secondary sequence within the numbered contig. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Alignment completed using CLUSTALW alignment program (http://www2.ebi.ac.uk/clustalw/). Boxshade server was used to block similar amino acid residues (http://www.ch.embnet.org/software/BOX_form.html).

A BLASTP search, with an increased E value from 10 to 10,000 was completed using this putative domain (data not shown). Since the region only spans seven amino acids, a large variety of proteins were identified, including two TOL-related proteins, but not TOL. Because of the length of the amino acid region, further analysis was not completed, as it is likely that a stretch of seven amino acids may occur randomly in a protein sequence and lead to misinterpretation of the "conserved" region.

2.4 DISCUSSION

2.4.1 Genomic comparison with TOL and *N. crassa* genomic contig sequences

Global database comparisons and motif search programs provide the opportunity to assign putative function to novel or newly identified proteins. With the completion of the *N. crassa* genome, a comparative base was established for identifying putative genes encoding homologous proteins within *N. crassa*. This approach was utilized to compare TOL and TOL-like sequences and to characterize and analyze the SET domain.

It was determined by TOL deletion analysis that amino acids position 99 to 521 are essential for TOL function, while amino acids in positions 652 to 790 are non-essential (Shiu and Glass, 1999; Shiu, 2000). The SET domain, defined from amino acids 336-515 (Smith *et al.*, 2000; Saupe *et al.*, 2000 - **Figure 5**), falls within the essential region. The region SET domain was identified by sequence similarity between TOL and HET-6 (OR and PA alleles) in *N. crassa* and HET-E in *P. anserina.* This similar region led to a search of the *N. crassa* genome to investigate if there were more "TOL-like" sequences present in the genome, with "TOL-like" sequences being defined as having significant amino acid sequence similarity with TOL (E < -1). Surprisingly, a large number of putative TOL-like sequences were identified. TOL BLASTP searches of the newly released *N. crassa* contig database from the Whitehead Institute revealed 42 potential TOL-like sequences.

2.4.2 Examination of the relevance of the SET domain

Examination of the TOL-like protein sequences in the *N. crassa* genome has showed that the SET region among all of the TOL-like sequences and HET-6 (OR and PA alleles) and HET-

E from *P. anserina* has been maintained (**Figure 14**). Comparing the entire pool of 42 TOL-like sequences, HET-6 (OR and PA alleles) and HET-E demonstrates a relatively high level of similarity in two of the three regions of the SET domain. Region one, using the defined region from this study (amino acid positions 343-351; **Figures 14 and 15a**), has 55.6% positivity (5/9; positivity = total number of identical and conservative or semi-conservative amino acid substitutions / total number of residues). Region two varies depending on which range of sequence is compared. Using the region spanning amino acid 377-433 in TOL (**Figure 14**), there is 14% positivity (8/57). This may not seem like a significant amount, but relative to entire sequence alignment, it is evident with the CLUSTALW consensus that this is a region with similarity. The third region of the SET domain is not significant in the overall alignment, with one evident semi-conservative amino acid substitution (**Figure 14**).

If only those sequences with high similarity in all three regions of the SET domain are compared with TOL, the pool of TOL-like sequences shrinks to nine (**Figure 23**). The similarity of the three regions in the SET domain increases to 66.7% positivity (6/9 for region 1), 45.6% positivity (26/57 for region 2) and 57.9% positivity (11/19 for region 3) (**Figures 15b, 17b, 20b** and **23**). There are a variety of possibilities in determining what makes the best alignment and which sequences are the most "TOL-like". This determination is subjective and decisions were made to maximize the number of identical amino acids and conservative amino acid substitutions. As in the results, the three regions in the SET domain have been redefined from spanning amino acid positions 336-382; 390-440 and 490-515 (combined from Smith *et al.*, 2000b and Saupe *et al.*, 2000) to 343-351, 377-433 and 491-509. In comparing the overall alignments (**Figure 14** versus **Figure 23**), all three regions span approximately the same amino acids. The noted difference is the change in conserved and similar amino acid residues per

region, reflecting the increased number of TOL-like sequences used in the alignments. The most significant difference is shown in the second region of the SET (**Figures 17a** and **b**) which was realigned by omitting the HET-6 (OR and PA alleles) and HET-E sequences in addition to other TOL-like contig sequences. The changes included the increase in the number of amino acids within region 2 from 39 to 57 and the increase in number of conserved amino acid residues from 6 to 7 and conservative and semi-conservative substitutions from 7 to 11.

Although the SET domain function remains unclear, BLASTP search results using individual regions with reference to TOL may provide some insight into potential functions for TOL and putative TOL-like proteins. A BLASTP of TOL (amino acids 343-509 [entire SET domain] and 377-433 [region two]; **Figures 18** and **24**) has similarity to Mms21p, a protein involved in DNA repair in *S. cerevisiae* (**Figure 19**). There is also similarity between the third region of the SET domain and the variable region of an immunoglobin kappa light chain protein, spanning 14 amino acid residues (**Figure 22**). This consensus includes TOL and 22 TOL-like sequences

2.4.2.1 Mms21p

There are only a few papers published that discuss the nature of Mms21p, which is characterized as a DNA repair protein in *S. cerevisiae*. It was first discovered as a recessive mutation in a screen to discover mutants that were sensitive to **m**ethyl **m**ethane**s**ulfonate (MMS), which is a monofunctional alkylating agent (Howard-Flanders and Theriot, 1966). *mms21-1* is located on chromosome V and is sensitive to UV and/or X-rays and has an increased rate of general spontaneous mutations and spontaneous mitotic recombinations as well as a decreased

efficiency of meiotic recombinations (Prakash and Prakash, 1977; Montelone *et al.*, 1981). *mms21-1* strains have a slower than wild type growth rate and appear to have defects in sporulation (Montelone *et al.*, 1981). The exact role of Mms21p is unclear and is stated as being involved in DNA repair due to epistatic relationships with *rad* genes, which are known components in DNA repair (referenced as unpublished data in Montelone *et al.*, 1981; Montelone and Koelliker, 1995). There is no definitive function for this protein and it remains as being generally described as a protein involved in DNA repair.

Mms21p and the second region of the SET domain in TOL have a stretch of 43 similar amino acid residues. A complete alignment of TOL and Mms21p contained many large gaps and did not show the similarity outside of the 43 amino acid region due to the large volume of sequence in the comparison (data not shown). Among the 43 amino acid residues, there are 12 identical amino acids and 16 conservative and three semi-conservative substitutions (**Figure 19**).

The significance of the similarity between TOL and Mms21p remains to be examined by experimental means, but since many DNA repair proteins influence or are influenced by DNA methylation of genomic sequence it may correlate methylation and TOL. Support for this argument has been shown in RIP analysis in a *tol-* background, where the efficiency of RIP is reduced (Selker, personal communications). As well, there is an increased recombination rate (40% - likely meiotic recombination) of *trp-4$^+$* and *tol^T/ili-1^T/isi^T* (Shiu, 2000), which may or may not correlate with the increased rate in spontaneous mitotic recombination in *mms21* mutants.

DNA repair in prokaryotic systems consists of four key pathways: DNA excision repair (Lindahl *et al.*, 1997), DNA double strand break repair (Karran, 2000; Featherstone and Jackson, 1999), mis-match repair (Modrich, 1991) and recombination repair (Oettinger, 1999). There are

many proteins involved in these pathways and they serve many catalytic, biochemical and structural roles. It may be that the sequence in the second region of the SET domain is present in proteins related to DNA repair function in *N. crassa* and potentially other organisms. There are a large number of **mu**tagen **s**ensitive mutants (*mus*) in *N. crassa* involved in excision repair, postreplication function and recombination (compiled in Perkins *et al.*, 2001). These genes are homologous with *rad* genes in *S. cerevisiae*. Future investigation could look at mutation rates of *tol-* strains after exposure to MMS, UV or X-rays. If there are increased rates of mutation, it may support that *tol* is involved in DNA repair. As well, double mutants with other DNA repair proteins may permit epistatic analysis.

The role of TOL in DNA repair and its relationship to vegetative incompatibility is a little hard to define, but there is characterized nuclear degradation associated with the cell death response. This DNA degradation requires proteins that may be involved in DNA repair and/or DNA metabolism. As well, some DNA repair mechanisms are coupled with methylation (i.e. mismatch repair in *Escherichia coli*; Modrich, 1991), either in recognition for repair or in methylation of the DNA itself, which may explain the alteration in RIP methylation patterns in *tol-* backgrounds (Selker, personal communications). There are indications that RIP and recombination may be correlated (Watters *et al.*, 1999), but direct evidence is still lacking.

2.4.2.2 Immunoglobin kappa light chain variable region

Immunoglobins are involved in the self/non-self recognition of foreign proteins in the mammalian system. There is a protein-protein recognition and interaction between antigens and antibodies, of which the antibodies are produced by the host as a defense and protection measure

against foreign particles. Antibodies have two chains, a heavy chain and a light chain, where variability and thus specificity can occur. As mentioned previously, the third region in the SET domain of TOL had similarity to the immunoglobin kappa light chain variable region (light chains exist as kappa or lambda, which have no functional differences – Becker and Deamer, 1991).

There are only 14 highly similar amino acids between the third region of the SET domain in TOL and the variable region of the immunoglobin kappa light chain. A BLASTP search with the third region in the SET domain (TOL-specific region) gave a positivity value of 70% with the variable region of the immunoglobin kappa light chain (positivity = number of identical amino acids and conservative substitutions / total number of amino acid residues; **Figure 21**). Of the similar amino acids, there were eight identical residues. However, a search of SCOP (Structural Classification of Proteins – data not shown; http://scop.berkeley.edu/) did not identify the region of TOL as having structural similarity to the variable regions in the immunoglobin kappa light chain. The significance of this region should therefore be viewed with caution. Variability in protein regions is often associated with specificity and recognition, in this case between antigens and antibodies. It is tantalizing to imagine this may be a region of variability between a number of proteins in *N. crassa* that function in the same means, as an immune system recognition. The large number of TOL-like sequences could then be explained as a mechanism for identifying self from non-self, through the complex formation and response to alternative protein products from the various *het* loci. This also fits nicely into the model of vegetative incompatibility and the presence of eleven *het* loci. TOL-like proteins could be the downstream amplification products for inducing vegetative incompatibility between the protein products of alternative *het* alleles (analogous to antigens). Specificity would permit precise TOL-like proteins to have function in

distinct *het* incompatibility reactions, like the antibody counterpart in response to the presence of foreign antigens.

2.4.3 Why are there so many TOL-like sequences in the *N. crassa* genome?

With the BLAST of TOL against the *N. crassa* genome, the question was raised, "why are there so many *tol*-like sequences in the *N. crassa* genome"? One possibility is that the TOL-like sequences serve as a component in regulation systems in maintaining control of vegetative incompatibility or possibly other functions within *N. crassa*, although there is little experimental evidence to indicate what these roles might be. A potential role in DNA repair is discussed in section **2.4.2.1**, but it may be that TOL and TOL-like proteins have function in other metabolic or homeostatic mechanisms in *N. crassa*.

Since there are so many TOL-like sequences identified, the SET domain could represent a novel domain. Many proteins are likely involved in the vegetative incompatibility response, with a number of different functions and domains - proteins that interact with DNA (i.e. zinc fingers or leucine zippers), interact with other proteins (i.e. leucine rich repeats or amphipathic α-helices) or with catalytic functions (i.e. kinases or methylases). One current theory for vegetative incompatibility involves convergence of *het* loci incompatibility pathways since there is a common downstream phenotype of hyphal compartmentation and death, in conjunction with a slowed growth rate and suppression of conidiation. This is a reproducible phenotype and may indicate common factors or proteins with similar functions. Support for this argument can be found in the examination of *vib-1*, a mutation that suppresses *het-c* vegetative incompatibility

and partially suppresses mating-type associated vegetative incompatibility (Q. Xiang and L. Glass, unpublished results).

There may also be some inherent redundancy because of the large number of TOL-like sequences, but it is only the SET domain that appears to be repeated, so this tends to support a more functional role. Unfortunately, all data comparisons at the present time are based on sequence similarity with no experimental evidence for the relevance of the domain. Future analysis may involve directed mutagenesis to examine the influence of mutations in the SET domain, using TOL as a model because it is molecularly characterized and has an identified role in mating-type associated vegetative incompatibility. *tol* is not an essential gene, as mutants show no obvious phenotype (Newmeyer, 1970; Shiu and Glass, 1999), and *tol* also appears to be non-functional in related species i.e. *N. tetrasperma*; Jacobson, 1992; Shiu, 2000).

2.4.4 What is the next step?

One area of future research would involve a more careful analysis of the TOL-like sequences, to determine if they are indeed in characterized ORFs. Upstream regions, start consensus sites, intronic sequences and termination signals would help support that the TOL-like sequences are regions of significance because it would increase the likelihood that they are true genes. The annotated *N. crassa* genomic sequence should be available at the end of the summer (2001), that would aid in this analysis.

As mentioned previously, another avenue of future research would be to do site directed mutagenesis of regions in the SET domain in TOL. Deletion analysis of TOL (Shiu and Glass, 1999; Shiu, 2000) found that the region containing the SET domain was essential for TOL

function. The second region of the SET domain, with similarity to a protein involved in DNA repair, Mms21p in *S. cerevisiae*, would be an appropriate area for research. This would permit evaluation of the significance of the SET domain in mediating the mating-type associated vegetative incompatibility response. Mutational analyses of the second region in the SET domain would also allow experiments to examine if there is increased mutation rates with exposure to MMS (**m**ethyl **m**ethane**s**ulfonate), UV or X-rays. Mutations in genes causing increased mutation rates with exposure to stimuli are associated with proteins involved in DNA repair, although the link of TOL and DNA repair is tenuous.

2.4.5 Conclusions

TOL contains a region, called the SET domain that is present in a large number of sequences in the *N. crassa* genome. BLASTP analyses have shown that a region in the SET domain has similarity to Mms21p, a protein involved in DNA repair in *S. cerevisiae*. The function for the SET domain remains unclear and future research may involve characterization of ORFs from the genomic sequence and subsequent cloning and mutational analyses to determine if there are phenotypic changes associated with the loss of the SET domain.

3. FINDING ADDITIONAL COMPONENTS IN THE MATING-TYPE ASSOCIATED VEGETATIVE INCOMPATIBILITY PATHWAY

3.1 INTRODUCTION

Mating-type associated vegetative incompatibility pathway has three identified components – MAT a-1, MAT A-1 and TOL. The purpose of this study was to identify additional components in the pathway. It was on the basis of the potential regions for protein-protein interactions in TOL that a yeast two-hybrid screen was performed with *tol* constructs. This established the basis for the research in this study, characterizing cDNAs encoding proteins that interact with TOL in the yeast two-hybrid system. The initial work was completed by P.K.-T. Shiu (Shiu, 2000) and is described below.

3.1.1 Yeast two-hybrid system

The yeast two-hybrid system exploits the Gal4p transcriptional activation system by separating the Gal4p DNA binding and transcriptional activation domains, both required for activation of gene transcription at an upstream activation sequence (UAS) in *Saccharomyces cerevisiae*. This allows *in vivo* examination of protein-protein interactions in an eukaryotic system, as proteins with the two domains must interact for transcription and translation of a reporter gene. Yeast cells are transformed with two constructs; a bait DNA sequence coding for a known protein is linked to the *GAL4* DNA binding domain sequence in one construct and fragments from a library or another known gene are linked to the transcriptional activation domain of *GAL4*. These constructs are then co-transformed into yeast for transcription and translation. If the bait and proteins encoded by cDNAs interact, the Gal4p DNA binding and

transcriptional activation domains come into close proximity and interact with the *GAL4* upstream activator sequence. This provides the means for transcription and translation of a downstream reporter gene such as *LACZ* or *HIS3*. *LACZ*, when transcribed and translated, produces β-galactosidase; when colonies are grown on media containing X-gal, the X-gal can be modified by β-galactosidase, which results in a blue colony. Negative clones do not produce β-galactosidase and therefore grow white in colour. *HIS3* serves as a nutritional marker. Colonies are grown on media that does not contain histidine, selecting for colonies expressing the *HIS3* reporter gene. If there is no interaction between the proteins of interest, there will be no expression of the corresponding reporter genes *LACZ* and *HIS3*. Finally, positive clones are verified through re-transformation back into yeast. For a visual explanation, refer to **Figure 26**.
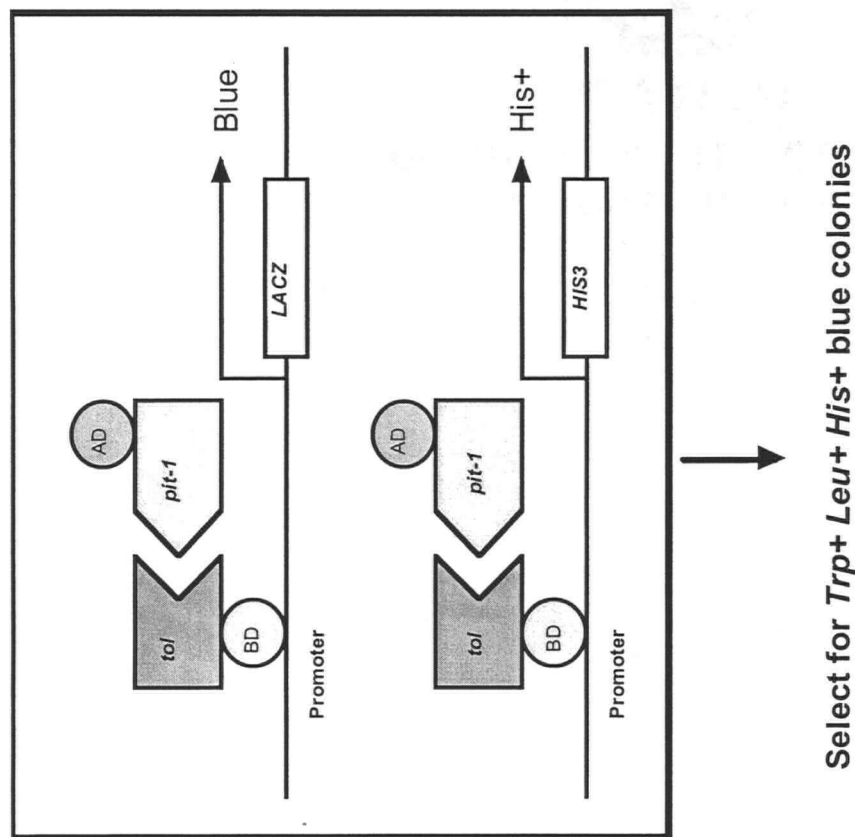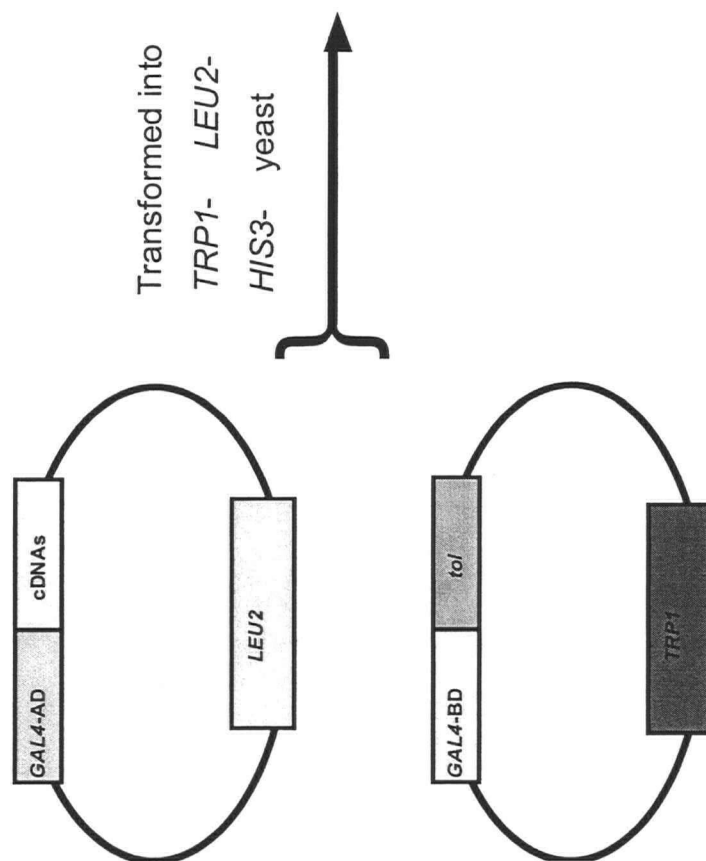
Figure 26. Yeast two-hybrid system

Figure 26. The yeast two-hybrid system. BD = DNA binding domain; AD = transcriptional activation domain (see text for details), *pit* genes (<u>p</u>roteins that <u>i</u>n<u>t</u>eract with <u>TOL</u>) – modified from a figure designed by P. K.-T. Shiu.

76

Reasons to use the yeast two-hybrid system include the potential to examine *in vivo* protein interactions and to investigate potentially weak interactions that would not survive stringent washing conditions of co-immunoprecipitation. This system also permits screening of a cDNA library for genes encoding putative proteins that interact with a protein of interest, i.e. TOL.

A major caveat with the yeast two-hybrid system is the prevalence of false positives, yeast colonies that express the reporter gene(s). The PCR process used to create the constructs may also incorporate mutations that void the relevance of a particular cDNA. A way of avoiding false positives is to utilize a combination of reporter genes, as in using *HIS3* and *LACZ* and to verify the constructs are in-frame by sequencing. Another disadvantage to using the yeast two-hybrid system is the lack of identification of cDNAs encoding proteins that do not manifest interactions in yeast or are transient. Some proteins need to be modified or activated in order to fully function and cooperate in protein-protein interactions or there may be incorrect folding or localization in the yeast background, leading to false positive or negative results.

3.1.2 Previous work

A Fungal Genetics Stock Center (FGSC) two-hybrid library made from FGSC 2489 mycelial cDNA was constructed in the HybriZap-2.1 vector system (Stratagene, La Jolla CA; website in **Figure 10**). This library was attached to the transcriptional activation domain of *GAL4*. The *GAL4* DNA binding domain vector was a tol/pGBT9 fusion construct, representing a partial *tol* sequence amplified using primers from -50 to -30 bp and ending at 1144 to 1129 bp (Shiu, 2000) linked to the *GAL4* DNA binding domain. The yeast strain used for the

transformation was YRG-2 (*Matα ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3112 gal4-542 gal80-538* LYS::UAS$_{GAL1}$-TATA$_{GAL1}$-HIS3 URA3::UAS$_{GAL4\ 17mers(x3)}$-TATA$_{CYC1}$-lacZ; Stratagene, La Jolla, CA). Reporter genes for this system were *HIS3* and *LACZ*. Thirty initially positive clones were sequenced in a single pass from the 5' end and re-transformed into yeast to verify the interaction and activity; five remained positive (**Table 1**, modified from Shiu, 2000).

Table 1. Positive TOL-interacting clones from a yeast two-hybrid screen

CLONE	CLONE SIZE (cDNA – bp)	AA SIZE	HOMOLOGY
7A5	700	> 46	Flavohemoprotein – *Alicaligens eutropus* (0.94)
7A10	800	> 106	Mucin-rat (medium similarity, 0.015)
7D3	1400	> 103	None in frame. Weak DNA similarity with a human clone (0.23)
7C10	950	> 201	None in frame. DNA similarity with chicken delta-1,2 (0.93)
7D8	1350	283	*vip1* (a p53-related protein) – *S. pombe* (7e-31)

Table 1. Modified Table 4-3 from Shiu, 2000. Positive clones identified from a yeast two-hybrid library screen using TOL as bait protein against a library of *N. crassa* mycelial cDNAs. Clone size refers to the size of cDNA insert in the pAD-GAL4 phagemid while AA size refers to number of amino acids encoded in-frame in the cDNA. Similarity with specific protein/DNA and their organisms of origin are given with p-value in bracket. The values in this chart are reproduced from Shiu (2000).

3.1.3 7D8/*ncvip1*

The 7D8 cDNA fragment was subcloned from the pAD-GAL4 construct (yeast two-hybrid *GAL4* activation domain vector; Stratagene, La Jolla CA) into the *Xho* I and *Eco*R I sites of pCB1004 (chloramphenicol and hygromycin-resistant phosphotransferase selection; Carroll *et al.*, 1994), completed by P.K.-T. Shiu (unpublished results). Protein sequence analysis of 7D8 (called *ncvip1* from this point on) showed similarity to Vip1 (GenBank accession number T37500; BLASTP E value of 7 e-31; **Figure 27**), a protein identified in a p53 expression library

screen of *S. pombe* using polyclonal anti-p53 antibodies. There is no published information on Vip1. A strain in which Vip1 was overexpressed showed sensitivity to cold temperatures and microtubule destabilizing drugs. Overexpression of a mutant form of *vip1* in *S. pombe* causes an alteration in cell shape and cell cycle arrest (A. Jungbluth, unpublished results). The nature of the similarity of Vip1 in *S. pombe* to human p53 is unknown (A. Jungbluth, personal communications). p53 is a tumor suppressor gene in humans, which acts as a transcriptional regulator; mutations in p53 are common in many cancers (Sheikh and Fornace, 2000).

Figure 27. Alignment of *N. crassa* NCVIP1 with *S. pombe* Vip1
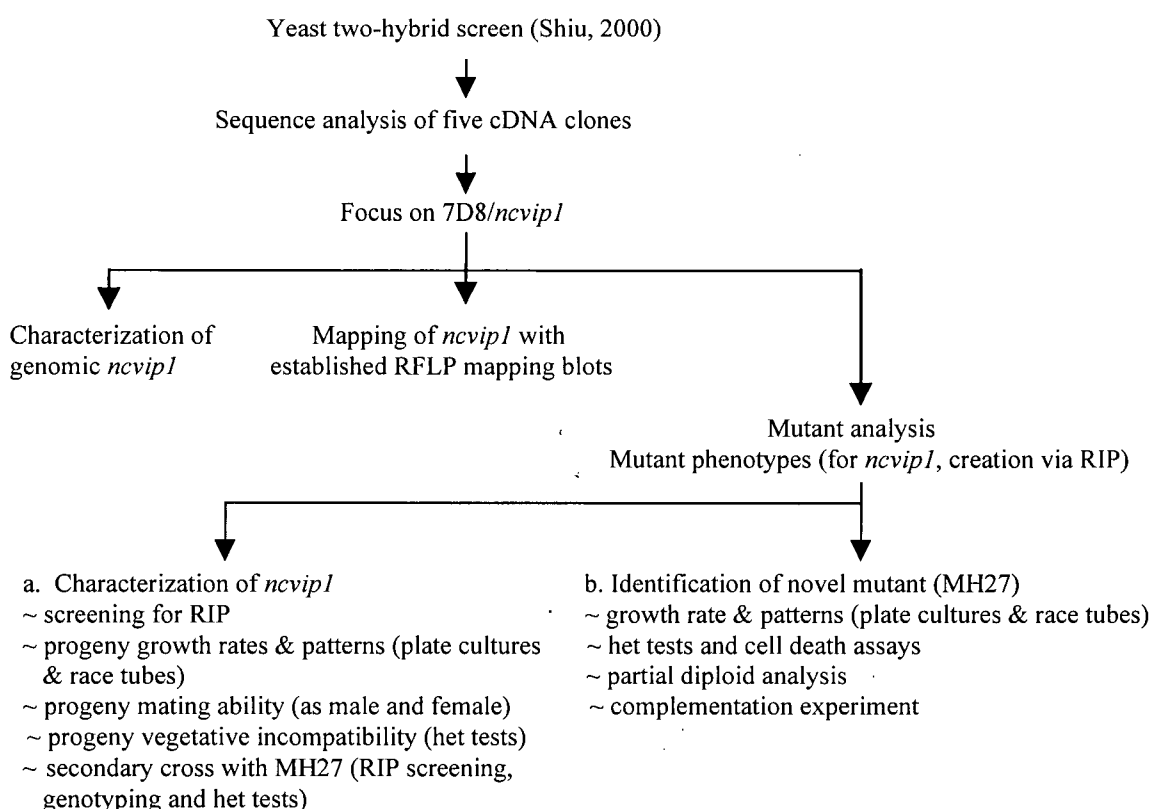
```
VIP1    MSNQVIVTNISPEVTEKQISDFFSFCGKVSNISTEKSGETQTAKIQFERPSATKTALLLQ 60
NCVIP1  MS-TVYVKNIGANTEEKDIRAFFSFCGKISSLDVTTEGETKSATVTFEKESAARTALLLD 59
        **   * *.**...:. **:*  *******:*.:.. ..***::*.: **: **::*****:


VIP1    DALLGQNKIQITSEDGGAASTTD----QGGAGGD-QAARQEDKPRSAIISELLSRGYHLS 115
NCVIP1  HTKLGEHELSVTSASGEHADSGDNVHPKSDADRDTDEITQEEKPRARVLAEYLASGYLVA 119
        .: **::::..:** .* *.: *     :...*. * :   **:***: :::* *: ** ::


VIP1    DVTLEKSIQLDQSYGVSSKFKGILESALSGVRSVNERYHVTEKANEVDNKFAISDKLNRT 175
NCVIP1  DSGLKTAIALDEKHGVSQRF-------LSTIQNLDQKYHATDRAKTADQSYGITARANSL 172
        *   *:..:* **:..:***.:*     ** ::..:::**.*::*: .*:..*: : *


VIP1    SSLVSTYFHKALETAAGTSAGQKVQNAYEDGKNQLLGIHNEARRLADAKNQAEG------ 229
NCVIP1  FSGLSSYFEKALEAPG----AKKIVDFYTTGSKQVQDIHNEAKRLAELKKQEAGGSSYKA 228
        * :*:**.****:..   .:*: : *  *..:*: .*****:***: *:*  *


VIP1    -------TASPASSTPTAPAEKEPTAPTTESKTTE------------------- 257
NCVIP1  AGLDKIFGAEKAPGQESKPNDQVPGAAPSDAAATESNQQPISEGAYPGTAEKIPQ 283
              *. *.. : * :: * *..::: :**
```

Figure 26. The CLUSTALW sequence alignment (http://www2.ebi.ac.uk/clustalw/) of putative NCVIP1 (modified sequence – see results) and Vip1 (GenBank access number: T37500). "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions.

79

ncvip1 homologues appear to be present in a large number of *Neurospora* species (*N. intermedia, N. discreta, N. sitophila, N. tetrasperma, N. terricola, N. dogei, N. galapagosensis, N. linolata, N. pannonica* and *N. africana* – Shiu, 2000). A blot containing samples of genomic DNA digested with *Eco*R I from the above species was probed with the *ncvip1*. Of these species, only *N. crassa* and *N. intermedia* exhibit mating-type associated vegetative incompatibility. The significance of the prevalence throughout *Neurospora* species is unclear.

3.2 APPROACH AND PROCEDURES: CHARACTERIZATION OF GENES PUTATIVELY INVOLVED IN MATING-TYPE ASSOCIATED VEGETATIVE INCOMPATIBILITY

Figure 28 - Outline for identification and characterization of genes putatively involved in mating - type associated vegetative incompatibility

Yeast two-hybrid screen (Shiu, 2000)

Sequence analysis of five cDNA clones

Focus on 7D8/*ncvip1*

Characterization of genomic *ncvip1*

Mapping of *ncvip1* with established RFLP mapping blots

Mutant analysis
Mutant phenotypes (for *ncvip1*, creation via RIP)

a. Characterization of *ncvip1*
~ screening for RIP
~ progeny growth rates & patterns (plate cultures & race tubes)
~ progeny mating ability (as male and female)
~ progeny vegetative incompatibility (het tests)
~ secondary cross with MH27 (RIP screening, genotyping and het tests)

b. Identification of novel mutant (MH27)
~ growth rate & patterns (plate cultures & race tubes)
~ het tests and cell death assays
~ partial diploid analysis
~ complementation experiment

To separate the two aspects of this study regarding potential components in the mating-type associated vegetative incompatibility pathway, *ncvip1* and MH27, "approach and procedures" and "results" will be divided into "a" and "b" sections respectively. This chapter finishes with a common discussion due to the overlap in the creation of the putative *ncvip1* RIP strains and the appearance of the novel mutation in MH27.

3a2.1 Additional analyses of putative *pit* genes from the yeast two-hybrid screen with TOL

3a.2.1.1 Plasmid construction and isolation

7A5, 7A10, 7C10 and 7D3 from the yeast two-hybrid screen were re-analyzed from the pAD-GAL4 vector constructs to verify putative positive clones. All constructs were maintained in a host *Escherichia coli* strain, DH5α (Hanahan, 1983; Life Technologies Inc., Rockville MD). Plasmid DNA was isolated via QIAprep Miniprep kits (Qiagen, Valencia CA); DNA was used for PCR and DNA sequence analysis. 7C10 was subcloned from pAD-GAL4 into pCB1004, using *Xba* I and *Eco*R I restriction sites. Since it was later determined that the 7C10 cDNA encoded a region too small for RIP analysis, subsequent transformation and RIP strain analyses were abandoned.

3a.2.1.2 DNA sequence analyses of putative *pit* genes

Sequencing of 7A10, 7C10 and 7D3 cDNA clones in the pAD-GAL4 vector was completed using vector specific T3 and T7 primers (**Table 2**); 7A5 was sequenced using pAD-GAL4 specific 5'AD and 3'AD primers (GAL4 Vector Recombinant Screening manual;

Stratagene, La Jolla CA - **Table 2**). *ncvip1* in pCB1004 was sequenced using *ncvip1* cDNA designed primers MDH1, MDH2, MDHS1 and MDHS2 (GibcoBRL Life Technologies Inc., Rockville MD; **Table 2**). Sequencing of two different clones in both directions was completed using the ABI automated Taq DyeDeoxy Terminator cycle method at UC Berkeley DNA Sequencing Facility - http://idrive.berkeley.edu/dnaseq/web. 7A5 was only sequenced once with 5'AD and 3'AD, as the cDNA insert was approximately 700 base pairs and so complete overlap was possible with the forward and reverse directions.

Computational analysis was completed using on-line software: BLAST (NCBI – website in **Figure 10**) and alignments using Multalin (website in **Figure 10**). Sequences were also compared to those previously identified by Shiu (2000).

Table 2. Primers for sequencing putative *pit* genes

PRIMER NAME	SEQUENCE (5'-3')	POSITION (bp) of *ncvip1* cDNA	TM (°C)
T7	GTA ATA CGA CTC ACT ATA GGG C		
T3	AAT TAA CCC TCA CTA AAG GG		
MDH1	ACA ACA GCA AAC ATT TCG G	1-19	57.16
MDH2	AAG ATT AGA GCA ACG CCA AC	1099-1080	57.12
MDHS1	TCA CTG GTT GAC GAC ACG	966-983	57.53
MDHS2	TTC TCC TCA GTG TTG GCG	129-112	58.41
5'AD	AGG GAT GTT TAA TAC CAC TAC		
3'AD	GCA CAG TTG AAG TGA ACT TGC		

Table 2. Primers used for sequencing putative *pit* genes. T7 and T3 are universal primers for amplification and sequencing of commercial vectors. MDH1, MDH2, MDHS1 and MDHS2 were designed from the *ncvip1* cDNA sequence using PRIMERS! (Ashland, MA). 5'AD and 3'AD sequences are provided through the GAL4 Vector Recombinant Screening manual (Stratagene, La Jolla CA). Position and TM for universal primers and 5'AD and 3'AD are not included.

3a.2.2 Characterization of *ncvip1*

3a.2.2.1 Strain maintenance, media and culturing

Table 3. Strains used in this study*

Spheroplasts and transformation strains	Genotype[a]	Source
201	*his-5 trp-4 tol; pan-2 a*	P.K.-T. Shiu
I-20-26	*ad-3B arg-1 A*	A.J.F. Griffiths
Crossing strains		
201	*his-5 trp-4 tol; pan-2 a*	P.K.-T. Shiu
FGSC 5448	*ad-3A; inl A*	FGSC[b]
FGCS 7214	*ad-8 lys-5 A*	FGSC[b]
I-20-41	*ad-3A arg-1; tol A*	A.J.F. Griffiths
Het testing strains		
R1-21	*pyr-4 arg-5 pregc-2 trp-3 a*	R.L. Metzenberg
I-20-41	*ad-3A arg-1; tol A*	A.J.F Griffiths
FGSC 5448	*ad-3A; inl A*	FGSC[b]
MH27.131	*ad-8 lys-5 A*	This study
I-1-51	*ad-3A nic-2 a*	A.J.F. Griffiths
C9-2	*thr-2 het6^{OR} hetc^{PA} a*	M. L. Smith
FGCS 7214	*ad-8 lys-5 A*	FGSC[b]
Xa-3	*arg-5 hetc^{PA}; pan-2 A*	Q. Xiang
Mating-type testers		
FGSC 4961	*f l a*	FGSC[b]
FGSC 7214	*ad-8 lys-5 A*	FGSC[b]
Helper strains		
FGSC 4564	*ad-3B cyh-1 a^{m1}*	FGSC[b]
New strains		
MH27	*pan-2 a*	This study

* All strains used in this study are Oak Ridge background, unless otherwise noted.

[a] *ad* = adenine; *arg* = arginine; *A*, *a* and *a^{m1}* = mating-type; *cyh* = cycloheximide resistant; *fl* = *fluffy*; *inl* = inositol; *het-6* and *het-c* = heterokaryon incompatibility loci; *his* = histidine; *lys* = lysine; *nic* = nicotinimide; *pan* = pathothenic acid; *pregc* = phosphatase regulation; *pyr* = pyrimidine; *trp* = tryptophan; *tol* = *tolerance*;

[b] FGSC = Fungal Genetics Stock Center; Department of Microbiology, University of Kansas Medical Center, Kansas City; website in **Figure 10**.

Strains used in the study of *ncvip1* are located in **Table 3**. Cultures of *N. crassa* were grown and maintained on minimal Vogel's medium slants (Vogel, 1964) with the appropriate

nutritional supplements at a 1x concentration. Conidial suspensions were stored at -80°C in a 15% sterile glycerol stock solution. Permanent stocks were established on silica gel (Brockman and deSerres, 1962).

Crosses were completed on Westergaard's medium (Westergaard and Mitchell, 1947), with the addition of pantothenic acid (D-pantothenic acid calcium salt; Fluka BioChemika; Sigma Chemical Co., St. Louis MO) when using parental strains containing a *pan-2* mutation. Forcing a heterokaryon between parental and a^{m1} strains (both OR background) allowed crosses to be completed without the addition of nutritional supplements, which can potentially interfere with the low nitrogen requiring condition for protoperithecial development. a^{m1} can not complete a cross, as it contains a base change mutation in *mat a-1* that renders the strain sterile (Griffiths and DeLange, 1978). Alternatively, crosses were also completed using 1/10X concentration of the required supplements, to ensure low levels of nitrogen. The strains used for mating assays were *fl a* (FGSC 4961) and FGSC 7214 + a^{m1} (**Table 3**).

3a.2.2.2 Genomic DNA isolation

Genomic DNA preparations were made from cultures grown in Vogel's liquid medium with the appropriate supplements at 1x concentration, following an extraction protocol by Metzenberg *et al.* (1988). This method is more laborious than others for genomic extraction from *N. crassa*, but a number of protocols were attempted and this method had the highest yield per mycelial material. The Metzenberg procedure uses a lysis buffer with triethylammonium salt (TEA) of EDTA (Sigma Chemical Co., St. Louis MO) and Pronase (Sigma Chemical Co., St. Louis MO). An ethanolic perchlorate solution is also used in a later step to remove proteins from

solution, while precipitating DNA (Metzenberg *et al.*, 1988). Genomic DNA was used in a number of experiments for PCR amplification and screening for RIP.

3a.2.2.3 PCR and DNA sequence analyses of *ncvip1*

ncvip1 was best amplified from the pCB1004 vector construct and genomic DNA samples using *ncvip1* cDNA-designed primers MDHF1 and MDH2 (GibcoBRL Life Technologies Inc., Rockville MD; **Table 4**). Polymerase chain reaction (PCR) amplifications were completed using the PTC-100 Programmable Thermal Controller (MJ Research Inc.). PCR products were cleaned either directly using the QIAquick Spin Columns (Qiagen, Valencia CA) or by cutting the appropriate band from a 0.8% agarose gel and purifying the DNA using the GeneClean kit (BIO 101, Carlsbad CA). Cleaned PCR products were subsequently cloned into pGEM (Promega, San Luis Obispo CA) or TOPO (Invitrogen, Carlsbad CA) PCR cloning vectors for sequence analysis. Vectors were maintained in *E. coli* strain, DH5α (Hanahan, 1983; Life Technologies Inc., Rockville MD).

Genomic *ncvip1* sequence was determined by sequencing the genomic cosmid G22:A12 (see section **3a.2.2.4** for cosmid identification) using *ncvip1* cDNA-designed primers MDH1, MDH2, MDHS1 and MDHS2 – GibcoBRL Life Technologies Inc., Rockville MD; **Table 4**, the wild type strain 201 and the mutant MH27 using MDHF1 and MDH2 *ncvip1* cDNA-designed primers – GibcoBRL Life Technologies Inc., Rockville MD; **Table 4**. Sequencing was completed using the ABI automated Taq DyeDeoxy Terminator cycle method at UC Berkeley DNA Sequencing Facility - http://idrive.berkeley.edu/dnaseq/web. 5' to 3' directions were sequenced for both DNA strands, as well as representatives from at least two different PCR

reactions to verify the correct sequence. Alignments were completed using Multalin (website in **Figure 10**) and CLUSTALW (website in **Figure 10**), comparing the established cDNA with the genomic *ncvip1* sequences. With the completion of the *N. crassa* genome (Whitehead Institute – website in **Figure 10**), the upstream region of *ncvip1* was available for examination. The ORF was analyzed using an ORF prediction program (http://genomic.sanger.ac.uk/gf/gf.shtml) and compared to *N. crassa* consensus characterization sequences published by Bruchez *et al.* (1993) and Edelmann and Staben, 1994.

Table 4. PCR amplification primers for *ncvip1*

PRIMER NAME	SEQUENCE (5'-3')	POSITION (bp) in *ncvip1* cDNA	TM (°C)
MDHF1	AGA TAC CAT CTC AAC CAC AGC	58-78	56.7
MDHR1	GAT CCA CCT CCA TTC TGC	967-950	56.31
MDH1	ACA ACA GCA AAC ATT TCG G	1-19	57.16
MDH2	AAG ATT AGA GCA ACG CCA AC	1099-1080	57.12
MDHS1	TCA CTG GTT GAC GAC ACG	966-983	57.53
MDHS2	TTC TCC TCA GTG TTG GCG	129-112	58.41

Table 4. Primers designed from the *ncvip1* cDNA used in PCR amplification of *ncvip1*. Names, sequence, position relative to the *ncvip1* cDNA and annealing temperature (TM) are listed. Primers were designed from the *ncvip1* cDNA sequence using PRIMERS! (Ashland, MA).

3a.2.2.4 Identification of a *ncvip1*-containing genomic cosmid

A set of blots containing *N. crassa* genomic strain 74ORSA Oak Ridge DNA (Orbach/Sachs pMCosX cosmid library – FGSC website listed in **Figure 10**) was probed for *ncvip1*, using a cDNA fragment labeled with 50μCi ^{32}P-dCTP (T7 QuickPrime kit - Amersham Pharmacia Biotech Inc., Piscataway NJ). Probes were cleaned using Quick spin Sephadex G columns for radiolabeled DNA purification (Roche Molecular Biochemicals, Indianapolis IN). The hybridization was completed according to Sambrook *et al.* (1989), using a hybridization

incubator (Robbins Scientific model 1000; Sunnyvale CA). Pattern visualization was completed using Kodak BioMax MS X-ray film (Eastman Kodak Company, Rochester NY) and autoradiography.

Genomic cosmids G22:A11 and G22:A12 were subjected to further analyses with single and multiple digests (*Bam*H I, *Cla* I, *Eco*R I, *Eco*R V, *Pst* I, *Bam*H I + *Cla* I, *Bam*H I + *Eco*R I, *Bam*H I + *Eco*R V and *Bam*H I + *Pst* I). These digests were probed with a *ncvip1* cDNA fragment. The Southern blot and hybridization were completed according to Sambrook *et al.* (1989). Probe labeling was achieved as described above. A control hybridization was performed with linearized pCB1004 vector.

3a.2.2.5 RFLP mapping of *ncvip1*

To determine the genomic location of *ncvip1*, it was possible to utilize a standard set of restriction fragment length polymorphism (RFLP) mapping strains. The strains represent a set of progeny created from two parental strains (Mauriceville-1c – a natural wild type strain - and Oak Ridge – a strain with known nutritional markers) that naturally have many polymorphisms (Metzenberg *et al.*, 1984; Perkins *et al.*, 2001; *N. crassa* genome website is listed in **Figure 10**). Thirty-eight recombinant progeny and the parental strains are run in a standard order on an agarose gel and used in a Southern hybridization with a genomic fragment probe containing a gene of interest. This generates an M/O (Mauriceville-like or Oak Ridge-like) restriction digest banding pattern, resulting from parental inheritance. Genes with known genetic location have been used as references (*N. crassa* genome homepage listed in **Figure 10**; Nelson *et al.*, 1998). The M/O sequence of a gene of interest is then compared to the known standards to determine

the genetic location. This mapping method is feasible for genes that have been cloned and have a corresponding genomic cosmid, since these larger probe fragments hybridize to a larger region and are likely to have a more diverse RFLP pattern. This method is in contrast to classical methods of mapping via percent of crossovers and recombination, which require a scorable phenotype.

For this study, a number of RFLP blots were hybridized with a genomic cosmid (G22:A12) containing *ncvip1*. The blot used to determine the overall RFLP M/O pattern was generated from progeny and parental strains cut with *Hin*d III. The probe was made by digesting G22:A12 cosmid DNA with *Xba* I, labeled and hybridized as mentioned previously. Visualization was completed using Kodak BioMax MS X-ray film (Eastman Kodak Company, Rochester NY), using the Konica medical film processor (Konica dealer, California Radiographics, Inc. Soquel CA).

3a.2.2.6 Repeat Induced Point (RIP) mutation analyses of *ncvip1*

3a.2.2.6.1 RIP background

RIP (Selker *et al.*, 1987; Selker, 1990 – **Figure 29**) is a transgene-induced gene silencing (TIGS) process that permits specific regional mutation targeting. It was thought to be unique to *N. crassa*, but was recently described in the creation of a nuclear migration mutant in *Podospora anserina* (Graïa *et al.*, 2001) and is identified as MIP in *Ascobolus immersus* (methylation induced premeiotically; Goyon and Faugeron, 1989; Rossignol and Faugeron, 1994). When there is duplicated sequence in the genome of *N. crassa*, the duplicated sequences are mutated when the strain is taken through a sexual cross. RIP can be exploited by introducing DNA from

a region of interest into *N. crassa* by transformation and using these strains in a sexual cross to selectively mutate the region of interest. The introduced DNA integrates in a random ectopic fashion, with one or more integrations (Fincham, 1989).

After fertilization, the transformed nuclei reside in the ascogenous hyphae, and by some unknown mechanism the duplicated sequences undergo G:C to A:T transitions in both copies of the duplicated region. It is postulated that pairing of the regions may be essential in recognition and identification of the duplicated sequences. A single copy integrand, which creates an ectopic and resident sequence pair, undergoes RIP more efficiently than a total of three or more copies (Fincham *et al.*, 1989). RIP occurs before karyogamy and meiosis. Cytosine methylation alterations are examined through genomic DNA digests with enzymes that are isoschizomers, such as *Dpn* II and *Sau*3A I. *Sau*3A I is methylation insensitive and is not blocked by *dam* methylation, whereas digestion by *Dpn* II is blocked by *dam* methylation. Hybridization to specific probes of DNA digested with these enzymes yields a differential banding pattern based on the methylation and restriction site differences and serve as a means to identify strains that have undergone the RIP process. The result of RIP is progeny that are mutated in the targeted area (**Figure 29**). This may create partial or complete loss-of-function mutants, depending on the degree of DNA alteration (Selker, 1990). The degree of RIPing depends on the gene and the size of insert.

RIP may serve as a protection measure from viral or foreign DNA that could integrate and disrupt functional genes in the genome. Recovery of RIPed progeny may be difficult if the gene is essential, but there are methods available for biasing towards nuclei containing RIP mutations in an essential gene (Metzenberg and Grotelueschen, 1992; Harkness *et al.*, 1994).

Figure 29. Repeat induced point (RIP) mutation analysis
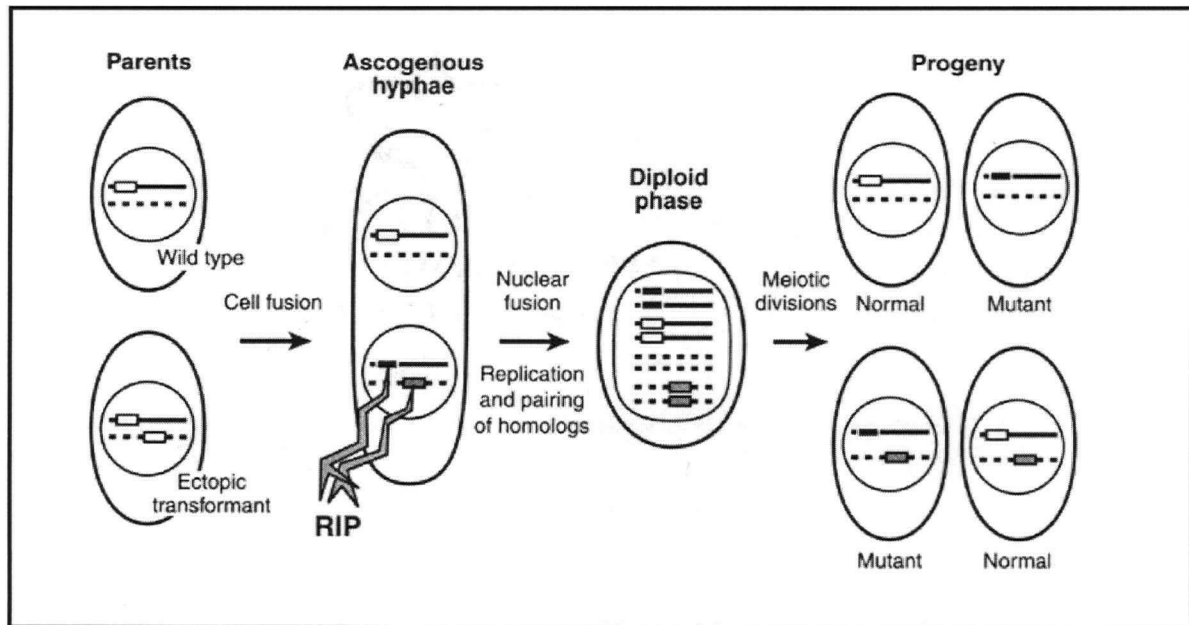
Figure 29. RIP analysis and progeny. A strain is transformed with a segment of DNA, creating an ectopic transformant. This strain is crossed with a wild type strain. After plasmogamy, but before karyogamy, the ectopic and resident copies in the ectopic transformant strain undergo G:C to A:T transitions and alter methylation patterns. The nuclei subsequently undergo karyogamy and meiosis to produce a linear tetrad of ascospores in which there is a 1:1 segregation of mutant to wild type nuclei. See text for additional details. Taken from Davis, 2000.

3a.2.2.6.2 Transformation of *N. crassa* with *ncvip1* for RIP

With the insertion of the *ncvip1* cDNA into pCB1004, *ncvip1* was available for manipulation and transformation into strains 201 (*his-5 trp-4 tol; pan-2 a*) and I-20-26 (*ad-3B arg-1 A*). The cDNA from *ncvip1* is 1.134 kb (excluding the poly A tail), which is more than the minimum requirement of 400 bp needed for RIP (Watters *et al.*, 1999). For spheroplast isolation, strain 201 was grown in liquid Vogel's minimal medium (Vogel, 1964) with 1x concentration of appropriate supplements. Spheroplasts were made according to Schweizer *et al.* (1981) and Vollmer and Yanofsky (1986). Strains transformed with pCB1004 vector and

pCB1004 derived constructs were grown on Vogel's medium slants, with 1x concentration of appropriate supplements and hygromycin B (CalBiochem, La Jolla CA) at 125 µg/ml.

Transformations of the *ncvip1* pCB1004 plasmid were completed using *tol⁻* (201 – *his-5 trp-4 tol; pan-2 a*) and *tol+* (I-20-26 – *ad-3B arg-1 A*) strains. The *tol-* and *tol+* backgrounds were used to avoid any complications from ectopic integration and potential overexpression of *ncvip1* in *mat A*, *mat a*, *tol+* or *tol⁻* backgrounds. Complications may also occur if there is a crucial interaction between the gene products of these genes for normal growth and function of *N. crassa*. I-20-26 transformations were conducted using electroporation of conidia (Margolin *et al.*, 1997 with modifications from R. Metzenberg, personal communications). One modification includes flaming a scoopula to collect conidia and prevent removal of agar, which can increase the salt concentrations and disrupt the electroporation process. The electroporation apparatus settings were altered to 25 µF capacitance, 600 Ω and 1.5 kV/0.1 cm. 201 transformations were completed using spheroplasts (made by P.K.-T. Shiu - Schweizer *et al.*, 1981; Vollmer and Yanofsky, 1986). Control transformations of pCB1004 into 201 and I-20-26 were also completed.

201and I-20-26 transformants were picked from the transformation plates containing 1% sorbitol in the top agar and 1X FIGS (20% L-sorbose, 0.5% D-fructose, 0.5% D-glucose, 2 ng/ml for 1L of a 10X stock) and 125 µg/ml hygromycin B in the bottom agar. Transformants were picked onto Vogel's minimal medium slants containing 1X concentration of the appropriate supplements and 125 µg/ml hygromycin B.

3a.2.2.6.3 Making homokaryons of *ncvip1*-transformed strains

To induce microconidial formation and ensure homokaryon cultures, I-20-26 (*ad-3B arg-1 A*) transformed strains were grown on IAA slants (Ebbole and Sachs, 1990; IAA = iodoacetic acid; Sigma Chemical Co., St. Louise MO). The resulting cultures were filtered with a Millipore 5 μm filter (Bedford MA) and selectively grown on BdeS plates containing 125 μg/ml hygromycin B (CalBiochem, La Jolla CA). BdeS (20% sorbose, 1% fructose and 1% sucrose in 1L for a 20X stock) is added to Vogel's salt medium (Vogel, 1964) at a final concentration of 1X to promote colonial growth. A select number of colonies was picked to Vogel's minimal medium slants, with 1x concentration of appropriate supplements and 125 μg/ml hygromycin B. 201 (*his-5 trp-4 tol; pan-2 a*) transformed strains did not grow on IAA medium and so these strains were directly plated onto BdeS plates containing 125 μg/ml hygromycin B for selection of individual colonies containing the transformation vector. This was repeated to bias towards selecting homokaryotic strains. Positive transformants were picked from the BdeS plates onto suitable Vogel's minimal medium slants with 1x concentration of the appropriate supplements and 125 μg/ml hygromycin B, maintaining selection for the pCB1004 vector.

3a.2.2.6.4 RIP crosses with *ncvip1*-transformed strains

The homokaryotic *ncvip1* transformation strains were crossed to a strain of the opposite mating-type to promote RIP mutations. 201 (*his-5 trp-4 tol; pan-2 a*) *ncvip1* and pCB1004 transformants were grown as the female and crossed to FGSC 5448 (*ad-3A; inl A*). I-20-26 (*ad-3B arg-1 A*) *ncvip1* and pCB1004 transformed strains were crossed to 201. Both crosses have the possibility of creating double mutants of *tol; ncvip1*. Strains were grown on Westergaard's

medium (Westergaard and Mitchell, 1947) and forced as a heterokaryon with a^{m1} (Griffiths and DeLange, 1978) to eliminate the need for the addition of supplements, with the exception of pantothenic acid. Strains used as male were grown on Vogel's minimal medium slants (Vogel, 1964) to induce conidia formation.

Crosses were completed with three homokaryotic strains of each parental type to provide a control comparison. The crosses were labeled as "MH" (201 [*his-3 trp-4 tol*; *pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 [*ad-3A*; *inl A*]), "B" (201/*vip*1.13.5 x FGSC 5448) and "C" (201/*vip*1.2.7.4 x FGSC 5448) for the three parental 201/*ncvip1* homokaryotic strains and "R" (201/PCB6.2 x FGSC 5448), "S" (201/PCB7.5 x FGSC 5448) and "T" (201/PCB2.3 x FGSC 5448) from the pCB1004 controls. Phenotypes caused by mutations in *ncvip1* should be consistent between all three 201/*ncvip1* crosses and not observed in the pCB1004 control cross progeny.

3a.2.2.6.5 Picking progeny from *ncvip1* RIP crosses

Ascospores from the RIP crosses were heat shocked in an aliquot of 150 µl of water at 65°C for 20-30 minutes and plated onto BdeS medium plates (recipe in section **3a.2.2.6.3**) for picking to Vogel's medium slants (Vogel, 1964), with 1x concentration of nutritional supplements. Progeny were picked from each of the three *ncvip1* 201 transformed parental crosses ("MH", "B" and "C" – crosses described in section **3a.2.2.6.4**) and from the pCB1004 201 transformed control crosses ("R", "S" and "T" – crosses described in section **3a.2.2.6.4**) using a dissecting microscope (Zeiss Stemi SV6; Zeiss, Thornwood NY). Progeny were screened for nutritional requirements by scoring for growth or lack of growth on minimal BdeS

medium plates versus growth on BdeS plates with 1x appropriate supplements (BdeS promotion of colonial growth permits assessment of genotype of multiple strains on the same plate). Hygromycin B resistance or sensitivity was also scored in the same manner, using BdeS medium plates with 125 μg/ml hygromycin B (CalBiochem, La Jolla CA). Favored progeny had nutritional requirements and were hygromycin sensitive, indicating the strains could be used in het tests and also to bias for the selection of RIPed resident copies of the gene of interest and against those with the pCB1004 vector.

3a.2.2.6.6 Assessing general growth and morphology of *ncvip1* RIP cross progeny

All progeny picked to Vogel's minimal medium slants (Vogel, 1964) with 1x concentration of appropriate supplements were assessed for general morphology and ability to produce conidia.

3a.2.2.6.7 Screening *ncvip1* RIP cross progeny for RIP

Genomic DNA isolations (see section **3a.2.2.2**) were completed with progeny having nutritional requirements and hygromycin sensitivity; progeny were examined from the *ncvip1* RIP "MH" and "C" crosses (two separate transformation strains of 201 [*his-5 trp-4 tol; pan-2 a*] with *ncvip1* crossed to FGSC 5448 [*ad-3A; inl A*]). RIP screening was established through digests of progeny genomic DNA with *Dpn* II and *Sau*3A I in two separate digests to identify methylation changes associated with RIP.

RIP screening involved two separate genomic digests with 3 µg DNA and 10 units

enzyme per µg of DNA, for both *Dpn* II and *Sau*3A I (New England Biolabs, Inc., Beverly MA)

in final reaction volumes of 15 µl. Digests were completed over a three-hour time period and

then loaded onto a 2X TAE, 1.5% agarose gel. Gels were run on low voltage (23-30 volts)

overnight in a 2X TAE buffer. The next day, gels were stained with ethidium bromide,

photographed under UV (Eagle Eye imaging system; Stratagene, La Jolla, CA) and blotted onto

Hybond-H+ nylon membrane (Amersham Pharmacia Biotech Inc., Piscataway NJ) according to

Sambrook *et al.* (1989). Transfers proceeded overnight. Gels were re-stained with ethidium

bromide and re-photographed after transfer to ensure the complete transfer of DNA to the

membrane. The probe, either PCR amplified *ncvip1* or linearized pCB1004, was then labeled

and cleaned, as described in section **3a.2.2.4**. The hybridization was completed according to

Sambrook *et al.* (1989). Visualization of methylation patterns was completed using Kodak

BioMax MS X-ray film (Eastman Kodak Company, Rochester NY) and autoradiography. In

total, 47 "MH" and 16 "C" progeny were screened and compared to controls of DNA from 201,

"MH", "C", "R" (crosses described in section **3a.2.2.6.4**) and FGSC 5448 (*ad-3A; inl A*).

3a.2.2.6.8 Heterokaryon (het) tests with the 17 "MH" progeny with methylation changes

Heterokaryon (het) tests (**Figure 30**) were used to examine vegetative incompatibility,

with respect to mating-type (Oak Ridge [OR] background). "MH" (201 [*his-5 trp-4 tol; pan-2 a*]

transformed with *ncvip1* and crossed to FGSC 5448 [*ad-3A; inl A*]) progeny with methylation

changes (MH26, MH27, MH38, MH'12, MH'16, MH'17, MH'36, MH'42, MH'45, MH'48,

MH'49, MH'52, MH'59, MH86, MH108, MH118, MH139 and MH146) were tested against

control Oak Ridge strains with *mat a* or *mat A* mating-type. Testers included either FGSC 5448 (*ad-3A; inl A*) or FGSC 7214 (*ad-8 lys-5 A*) as the *mat A* tester and R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*) and I-1-51 (*ad-3A nic-2 a*) as the *mat a* tester and control. Control tests were completed between *mat a* + *mat a* (R1.21 + I-1-51 = compatible), *mat A* + *mat A* (FGSC 5448 + FGSC 7214 = compatible) and *mat a* + *mat A* (FGSC 5448 + R1.21 or FGSC 7214 + R1.21 = incompatible) for comparison. Conidial suspensions of strains were co-inoculated onto Vogel's minimal medium plates (Vogel, 1964) and growth marked at 24-hour intervals.

Figure 30. Heterokaryon (het) tests

A (*thr-2)*   B (*his-5*)

Minimal Medium plate
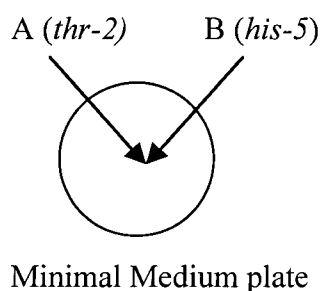
Figure 30. Heterokaryon test. Conidial suspensions are co-inoculated onto Vogel's minimal medium plates (Vogel, 1964). In order to grow, the strains must form a heterokaryon to complement nutritional deficiencies. Vegetative incompatibility is determined when there is a failure to produce a wild type growth and conidiation pattern in the heterokaryon. There is also an increase in cell death in the heterokaryon if it is incompatible. As mentioned in section **1.4**, vegetative incompatibility is mediated by genetic differences at 11 different *het* loci in *N. crassa.*

3a.2.2.6.9 Growth rate analyses of "MH" progeny

Growth rates using race tubes (Ryan *et al.*, 1943) were determined for six "MH" progeny (201 [*his-5 trp-4 tol; pan-2 a*] transformed with *ncvip1* crossed to FGSC 5448 [*ad-3A; inl A*]; MH26 [*his-5 trp-4 tol A*], MH27 [*pan-2 a*], MH38 [*his-5; pan-2 A*], MH'12 [*pan-2; inl A*], MH'16 [*his-5 trp-4 tol; pan-2 A*] and MH'17 [*pan-2 A*]), completed in triplicate at room temperature and 10°C. Comparison of growth rate was made with respect to control strains, 201

(*his-5 trp-4 tol; pan-2 a*) and FGSC 5448 (*ad-3A; inl A*). Cultures were grown on minimal

Vogel's medium (Vogel, 1964) with 1x concentration of appropriate supplements. Growth was

marked at 24-hour intervals and data was analyzed using Microsoft Excel to determine an

average growth per day and the appropriate standard deviations.

3a.2.2.6.10 Reproductive ability of the 17 "MH" progeny with methylation changes

Sexual development of the 17 "MH" progeny (201 [*his-5 trp-4 tol; pan-2 a*] transformed

with *ncvip1* crossed to FGSC 5448 [*ad-3A; inl A*]) with potential mutations in *ncvip1* was

assessed through the strains' ability to complete a cross as a female (on Westergaard's medium –

Westergaard and Mitchell, 1947), with the addition of a wild type male partner. Strains were

also tested for mating-type via conidial suspensions, thus assessing the ability to complete a

cross as a male parent. In all, these tests determined the ability of a strain to complete the sexual

cycle.

3a.2.2.6.11 Secondary cross with MH27 to segregate the methylation changes associated with the RIP of *ncvip1* from the pCB1004 vector

Despite selection of hygromycin sensitive progeny, the 17 strains identified as having

methylation changes associated with the RIP of *ncvip1* also hybridized to a vector pCB1004

control probe. Therefore, a secondary cross with MH27 (*pan-2 a* – female parent) and FGSC

7214 (*ad-8 lys-5 A* – male parent) was completed to segregate the methylation changes from the

vector. Crosses were completed on Westergaard's medium plates (Westergaard and Mitchell,

1947), with the addition of pantothenic acid.

3a.2.2.6.12 Picking progeny and genotype assessment of MH27 derived strains

Ascospores were heat shocked in an aliquot of 150 μl of water at 65°C for 20-30 minutes and plated onto BdeS plates (recipe in section **3a.2.2.6.3**) for picking to Vogel's medium slants (Vogel, 1964) with 1x concentration of nutritional supplements. Picking conditions are discussed in section **3a.2.2.6.5**. Progeny were screened for nutritional requirements using BdeS medium plates, as mentioned previously. Hygromycin sensitivity was not assessed, as the original MH27 strain was hygromycin sensitive. Mating-type was assessed using Westergaards medium (Westergaard and Mitchell, 1947) and tester strains *fl a* (*mat a*) and 7214 + a^{m1} (*mat A*) as female parents.

3a.2.2.6.13 Screening MH27 RIP cross progeny for RIP

RIP screening was completed as previously mentioned (section **3a.2.2.6.7**), using genomic DNA digested with *Dpn* II and *Sau*3A I (New England Biolabs, Inc., Beverly MA). Forty-three MH27 derived progeny were screened, with the Southern blot and hybridization conditions as mentioned previously (section **3a.2.2.6.7**). One modification was in the visualization of the hybridizations. Storage Phosphor screen cassettes from Molecular Dynamics (Amersham Pharmacia Biotech, Piscataway NJ) were used to capture the emitted β-particles and assessed using the Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Pharmacia Biotech Inc., Piscataway NJ) with the associated software. Images were manipulated using Adobe Photoshop 5.5 (Adobe Systems Inc., San Jose CA). When the Phosphor Imager was unavailable, visualization was accomplished using Kodak BioMax MS X-ray film (Eastman Kodak Company, Rochester NY).

3b.2.1 Identification of a novel mutation influencing the mating-type associated vegetative incompatibility response

As mentioned previously, 17 "MH" (201 [*his-3 trp-4 tol*; *pan-2 a*] transformed with *ncvip1* x FGSC 5448 [*ad-3A*; *inl A*]) *ncvip1* RIP cross progeny with methylation changes were screened for influence on mating-type associated vegetative incompatibility via het tests (section **3a.2.2.6.8**). One of these progeny, MH27, showed a partial suppression of mating-type associated vegetative incompatibility (designated the "Sup (mvi)" phenotype for <u>sup</u>pressor of <u>m</u>ating-type associated <u>v</u>egetative <u>i</u>ncompatibility). As described in section **3a.2.2.3**, *ncvip1* from MH27 was sequenced. It was determined that MH27 had an intact copy of *ncvip1*, indicating a novel phenotype associated with this strain. MH27 was thus subject to further analysis. A cross between MH27 (*pan-2 a*) and FGSC 7214 (*ad-8 lys-5 A*) is described in section **3a.2.2.6.11**.

3b.2.1.1 Growth rates of MH27 derived progeny

MH27.11 (*ad-8 lys-5 a*), MH27.14 (*ad-8 lys-5 a*) and MH27.131 (*ad-8 lys-5 A*) from the cross of MH27 and FGSC 7214 (described above) were used to assess the amount of suppression of mating-type incompatibility in race tubes compared with the MH27 strain, using Vogel's minimal medium (Vogel, 1964) with 1x concentration of the appropriate supplements. Race tubes (Ryan *et al.*, 1943; completed in triplicate for statistical purposes) were completed with MH27, MH27.14 and MH27.131 at room temperature and 10°C to assess potential influences of temperature on growth, with growth being marked at 24-hour intervals. MH27.11 was assessed at room temperature only. Results were analyzed using Microsoft Excel by determining an average growth rate per day and calculating standard deviations for variance.

3b.2.1.2 Heterokaryon (het) tests for mating-type and *het-c* vegetative incompatibility with MH27 derived progeny

Genotype assessment of MH27 progeny was described in section **3a.2.2.6.12**. Sixty MH27 progeny were screened to determine the heredity of the Sup (mvi) phenotype. Again, the *mat A* tester was FGSC 5448 (*ad-3A; inl A*) and the *mat a* tester was R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*). Tests were completed by co-inoculating a conidial suspension of each parental strain onto Vogel's minimal medium plates (Vogel, 1964). Growth was noted at 24, 48 and 72 hours after inoculation.

The same progeny assessed for growth rates (MH27.11 [*ad-8 lys-5 a*], MH27.14 [*ad-8 lys-5 a*] and MH27.131 [*ad-8 lys-5 A*]) were also grown in race tubes (Ryan *et al.*, 1943), using Vogel's minimal medium to determine the growth rates. Again, assessments were made at room temperature and 10° C (with the exception of MH27.11) and marked at 24-hour intervals. Data were analyzed using Microsoft Excel to determine the average growth rate per day and the corresponding standard deviations.

Influence of the Sup (mvi) phenotype was assessed with respect to *het-c* vegetative incompatibility. Het tests were prepared between MH27 (*pan-2 a*), MH27.11 (*ad-8 lys-5 a*), MH27.14 (*ad-8 lys-5 a*), MH27.69 (*pan-2 A*) and MH27.173 (*ad-8 lys-5 pan-2 a*) with *het-c* testers C9-2 (Oak Ridge - *thr-2 het-c^{PA} a*) and Xa-3 (Panama - *het-c^{PA} arg-5; pan-2 A*). Appropriate mating-type and *het-c* het test controls with C9-2 and Xa-3 were completed for comparison. Tests were performed by co-inoculating conidial suspensions on Vogel's minimal medium plates and growth noted after 24, 48 and 72 hours.

3b.2.1.3 Assessing vegetative and sexual growth and development of MH27 derived progeny

Conidiation levels (male structures) were qualitatively assessed for all MH27 derived progeny by observing growth on Vogel's medium slants (Vogel, 1964), with 1x concentration of appropriate supplements. Mating-type of the MH27 derived strains was assessed against *fl a* (*mat a*) and FGSC 7214 + a^{m1} (*mat A*) testers. These tests also evaluated the strains' ability to complete a sexual cross as a male parent.

Sexual development was assessed through the strains' ability to produce female structures in inducing conditions (Westergaard's medium; Westergaard and Mitchell, 1947), and the ability of a strain to complete the sexual cycle as a female. This was completed for MH27.11 (*ad-8 lys-5 a*), MH27.14 (*ad-8 lys-5 a*), MH27.69 (*pan-2 A*) and MH27.131 (*ad-8 lys-5 A*). Strains were grown on Westergaard's medium with 1/10x concentration of appropriate supplements to ensure low nitrogen conditions.

3b.2.1.4 Assessing cell death

Cell death was assayed in het tests between MH27 (*pan-2 a*), MH27.11 (*ad-8 lys-5 a*) and MH27.14 (*ad-8 lys-5 a*) with R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*; compatible tester) and MH27.131 (*ad-8 lys-5 A*; incompatible tester). Compatible and incompatible controls were also assayed, as were the death rates in the individual strains. Het tests were completed on Vogel's minimal medium plates (Vogel, 1964), with a strip of sterile cellophane to permit mycelial manipulation and staining. Individual strains were grown on complete Vogel's medium, with the strip of sterile cellophane for mycelial cell death assessment.

Evan's Blue (Direct Blue 53 - Sigma Chemical Co., St. Louise MO) is the dye used to assess cell death, as it is a vital dye that is excluded from intact and living cells (Gaff and Okong'O-Ogola, 1971). The mycelial material on the cellophane strip was removed from the plate using sterile forceps and stained in 1% Evan's Blue for five minutes. This was followed by a five-minute rinse with sterile, distilled water to remove excess stain. The mycelia were placed on a glass slide and observed using bright field microscopy (Nikon Optiphot-2). Cell death was determined by counting the number of live versus dead cells in a sample, with a total cell count of approximately 200 cells. From these counts, the percentage of cell death was calculated. For this study, only one count was completed for each individual strain and het test condition.
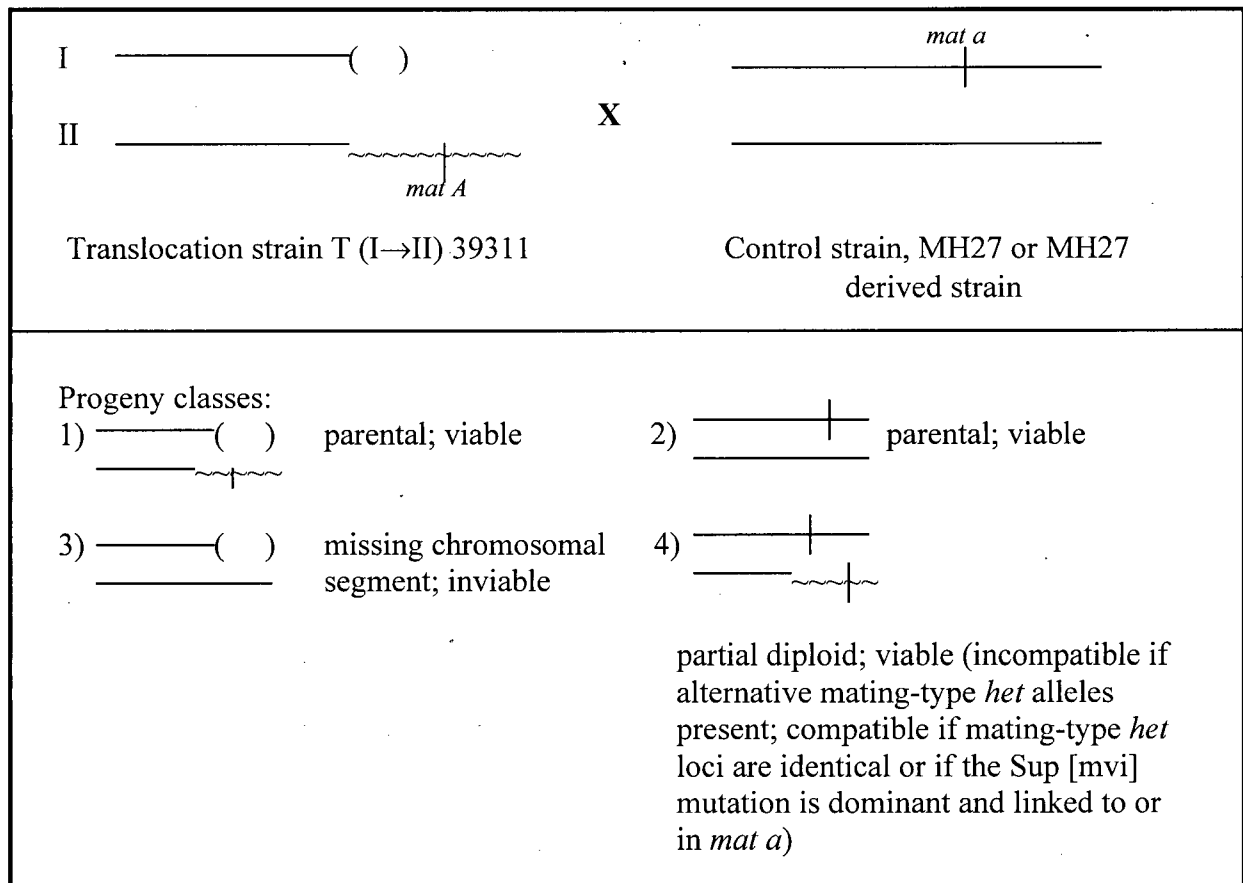
3b.2.1.5 Partial diploid analyses

3b.2.1.5.1 Background on partial diploid analysis

Partial diploid analysis has been used to generate duplications in portions of the genome in *N. crassa* (**Figure 31**). Perkins (1975) and Mylyk (1975) used translocation and inversion strains to identify a large number of *het* loci in *N. crassa*. *het* loci were screened by the generation of an incompatible phenotype, resulting from the presence of alternative alleles at a *het* locus because of a regional duplication. Partial diploid analysis involves crossing a normal sequence strain to a strain with a translocation or an inversion; explanations will be made using a translocation strain since this was the nature of the strain used to generate partial diploids in this study (**Figure 31**). The segregation of the two linkage groups concerning the translocation (linkage groups I and II in this study) and pertaining to the region of partial duplication, results in four classes of progeny. Two classes represent the parental arrangement of translocation linkage

groups, while the third represents a region of partial duplication. The fourth class of progeny is inviable because part of one linkage group is missing (**Figure 31**). The partial diploid class of progeny is of most interest in examining vegetative incompatibility, *het* loci and potential suppressors.

Figure 31. Partial diploid analysis

Figure 31. Aneuploid or partial diploid analysis permits examination of duplicated region within the same nucleus. Translocation strains, T (I→II) 39311 *mat A,* were crossed to a normal sequence strain (in this study, I-1-51, MH27, MH27.14 and MH27.45). This permits examination of the Sup (mvi) phenotype to determine if it suppresses the incompatible phenotype associated with partial diploid progeny (class 4), at the *mat* locus. If the Sup (mvi) mutation does not suppress the *mat A/ mat a* incompatibility, a 2:1 segregation of compatible to incompatible progeny is predicted among the viable progeny.

3b.2.1.5.2 Partial diploid analyses with MH27, MH27.14 and MH27.45

MH27 (*pan-2 a*) and two of its progeny (MH27.14 [*ad-8 lys-5 a*] and MH27.45 [*pan-2 a*]) were crossed to translocation strains to generate partial diploid progeny for the *mat* region. Partial diploid analysis was completed by crosses with R4.33 (*T (IL→IIR)* 39311; *trp-4 tol; A*) and R4.35 (*T (IL→ IIR)* 39311; *ser-3 A*); I-1-51 (*ad-3A nic-2 a*) served as a wild type *mat a* control for crosses with the translocation strains. Crosses were named as follows:

PDA = MH27.14 x R4.35   PDE = MH27.45 x R4.35

PDB = MH27.14 x R4.33   PDF = MH27.45 x R4.33

PDC = MH27 x R4.35   PDI = I-1-51 x R4.35

PDD = MH27 x R4.33   PDJ = I-1-51 x R4.33

PDA, PDB, PDC, PDD, PDE, PDF, PDH and PDI ascospores were heat shocked in an aliquot of 150 µl water at 60-65°C for 25 minutes and picked to Vogel's minimal medium slants (Vogel, 1964) with 1x concentration of appropriate supplements. Viability and general growth of progeny were noted. Progeny were also screened for genotype using BdeS plates (recipe in section **3a.2.2.6.3**) and mating-type with *fl a* (*mat a*) and 7214 + a^{m1} (*mat A*), as mentioned previously.

3b.2.1.6 MH27 complementation of the Sup (mvi) phenotype: transformations and het tests

3b.2.1.6.1 Plasmid construction and transformation into MH27 and a^{m1}

mat a-1 was subcloned from pCSN4 (ampicillin selection – Staben and Yanofsky, 1990) into the pCB1004 transformation vector using *Xho*I and *Eco*RI restriction sites. A *mat A-1* clone in pCB1004 (11.2.95.3 – Shiu, unpublished construct) was used as a comparative control for transformation assays into a^{m1} (*ad-3B cyh-1 a^{m1}*) and MH27 (*pan-2 a*). A host *Escherichia coli* strain, DH5α (Hanahan, 1983; Life Technologies Inc., Rockville MD), was used for plasmid maintenance and amplification.

MH27 and a^{m1} spheroplasts were made according to Schweizer *et al.* (1981) and Vollmer and Yanofsky (1986), grown in liquid Vogel's minimal medium (Vogel, 1964) with 1x concentration of appropriate supplements. Modifications were made in the enzyme use; 10 mg/ml Glucanex (Novo Nordisk Ferment Ltd., Dittingen Sweden) and 50 µg/ml driselase (Sigma Chemical Co., St. Louise MO) were used in place of Novozyme. MH27 and a^{m1} spheroplasts were transformed with pCB1004, *mat a-1* and *mat A-1* both in the pCB1004. Transformants were picked from the transformation plates to Vogel's minimal medium slants, containing 1x concentration of the appropriate supplements and 200 µg/ml hygromycin B (CalBiochem; La Jolla CA).

3b.2.1.6.2 Creating homokaryotic MH27 and a^{m1} transformant strains

As mentioned in section **3a.2.2.6.3**, MH27 (*pan-2 a*) and a^{m1} (*ad-3B cyh-1 a^{m1}*) transformants were grown on IAA medium (Ebbole and Sachs, 1990) to induce microconidial formation. The cultures filtered through a 5 µm Millipore filter (Bedford MA) and plated onto

BdeS plates (recipe in section **3a.2.2.6.3**), containing 1x concentration of appropriate supplements and 200 μg/ml hygromycin B (CalBiochem; La Jolla CA). Plates were incubated at 30°C for four days and transformants picked to Vogel's slants with 1x concentration of appropriate supplements and 200 μg/ml hygromycin B.

3b.2.1.6.3 Heterokaryon tests with MH27 and a^{m1} transformation strains

Het tests were completed with at least three MH27 and a^{m1} transformants, each with pCB1004, *mat a-1* or *mat A-1*. Control het tests were completed using R1.21 + I-1-51 (*pyr-4 arg-5 pregc-2 trp-3 a* + *ad-3A nic-2 a* = compatible), MH27.131 + FGSC 5448 (*ad-8 lys-5 A* + *ad-3A; inl A* = compatible) and MH27.131 + R1.21 (incompatible; genotypes above). Experimental tests were completed using short race tubes (test tube length) and tested against R1.21 and MH27.131. Growth was marked at 24-hour intervals. Mating-type of the heterokaryons was tested against *fl a* (*mat a*) and FGSC 7214 + a^{m1} (*mat A*) tester strains.

3.3 RESULTS

As mentioned in the beginning of the Approach and Procedures section, the results will be similarly divided into "a" and "b" sections for *ncvip1* and MH27 and the Sup (mvi) phenotype respectively.

3a.3.1 Sequence analyses of putative *pit* genes from the yeast two-hybrid screen with TOL

tol, a recessive suppressor of mating-type associated vegetative incompatibility, has an amphipathic α-helix and a leucine rich repeat that may be involved in protein-protein interactions. To examine genes that potentially encode proteins that interact with TOL, TOL was used as bait in a yeast-two hybrid screen.

Shiu (2000) identified five cDNA-pGAL4 plasmid dependant clones that encode protein products that interacted with TOL in the yeast two-hybrid system. These clones were re-sequenced and compared with the 5' initial single pass sequence (**Table 5** and **Appendix**); 7A5 was deemed as a false positive, since it only encodes two amino acids. 7C10 and 7D3 encode 74 and 56 amino acids respectively. Putative protein-protein interaction domains may or may not be represented in these short regions, but it is likely that the cDNAs would be insufficient for efficient RIP analysis. The remaining two clones, 7A10 and 7D8 (called *ncvip1* from this point on), yielded putative polypeptides of 130 and 283 amino acids respectively. Due to a previously identified homology with *vip1* in *S. pombe*, this project thus focused on *ncvip1*.

Table 5. cDNAs encoding putative TOL-interacting proteins

CLONE	STATUS	SIZE OF PUTATIVE PROTEIN REGION (aa)	
		INITIAL PREDICTION	NEW PREDICTION
7A5	False positive	>46	2
7A10	Putative positive	>106	130
7C10	Putative positive*	>201	74
7D3	Putative positive*	>103	56
7D8 (*ncvip1*)	Putative positive	283	283

*Note: False and putative positives are relative terms. False positive refers to a polypeptide region too small for a true interaction. Putative positives (*) are cDNAs that demonstrate a positive interaction with TOL in the yeast 2-hybrid system, but the polypeptide regions are too small to be analyzed efficiently through RIP analysis. Therefore, they are not a focus of this project.

NCBI GenBank database BLASTP searches were completed with 7A10, 7C10, 7D3 and 7D8 polypeptides (data not shown). A BLASTP search with 7A10 had a number of hits, but only two with E values less than one (0.18 and 0.57 for two hypothetical proteins in *Arabadopsis thaliana*). No protein homologues had an identity or positivity value above 50% (identity = identical amino acid residues / total number of residues x 100%; positivity = identical + conservative amino acid residues / total number of residues x 100%). Similar results were determined for 7C10 and 7D3. Protein motif analyses of 7A10, 7C10 and 7D3 did not yield any significant domains.

BLASTP was already completed with the NCVIP1, as described in section **3.1.3**. Output results find one significant similar sequence, gi|7493708|pir||T37500 Vip1 from *S. pombe* – E value = 7 e-31. Six unidentified proteins from *A. thaliana* were identified with E values ranging from 3 e-05 to 6 e-09, but the relative identity and positivity values were under 50% (data not shown). Because of the large number of unidentified or putative proteins and lack of proteins with identifiable function distinguished in the NCVIP1 BLASTP search, the results are not included.

3a.3.2 7D8/*ncvip1*

3a.3.2.1 Sequence characterization of *ncvip1*

### 3a.3.2.1.1 Identification and characterization of a genomic cosmid containing *ncvip1*

To further characterize the genomic sequence of 7D8 (*ncvip1* from this point on), the *ncvip1* cDNA was used to probe a *N. crassa* genomic library (Orbach/Sachs pMCosX cosmid library), identifying cosmid G22:A12. To characterize a smaller *ncvip1*-containing fragment of G22:A12, G22:A12 and a control cosmid G22:A11 (near G22:A12 on the blot, but not containing *ncvip1*) were digested with a number of enzymes (**Figures 32a and b**). Blots of the digested DNA were probed with *ncvip1* in the pCB1004 vector and with the pCB1004 vector alone in a second hybridization. There was a common 4 kb band in *Bam*H I and *Bam*H I combination digests of G22:A12 using both *ncvip1* and the pCB1004 probes, indicating a doublet in which one 4 kb band contained *ncvip1* and the other a portion of the vector. The 4kb band was missing in the *Bam*H I + *Pst* I digest hybridized to pCB1004, but present when probed with *ncvip1*, showing the reduction of the doublet to one 4 kb band, in which *ncvip1* was present. Attempts to subclone the 4 kb *ncvip1*-containing *Bam*H I + *Pst* I fragment from G22:A12 into pCB1004 were unsuccessful.

Figure 32a. G22:A11 and G22:A12 genomic cosmid digests hybridized to *ncvip1*
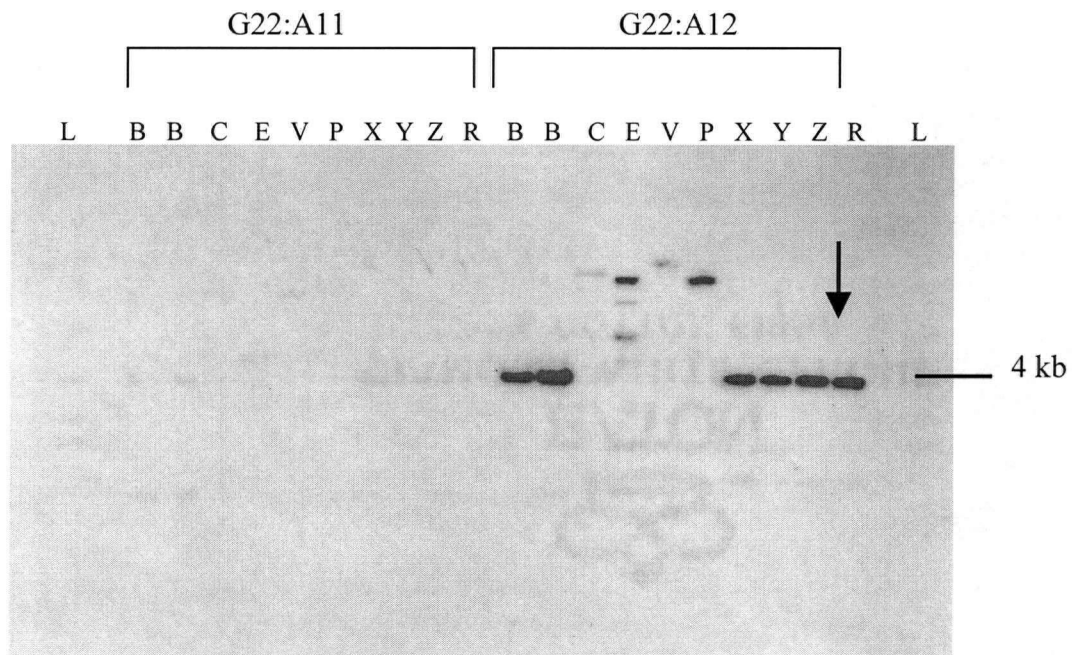
Figure 32a. G22:A11 and G22:A12 genomic cosmids from the Orbach/Sachs library probed with PCR amplified *ncvip1*. Blot exposed for 15 minutes on X-ray film (autoradiography). L= ladder, B= *Bam*H I, C= *Cla* I, E= *Eco*R I, V= *Eco*R V, P= *Pst* I, X= *Bam*H I + *Cla* I, Y= *Bam*H I + *Eco*R I, Z= *Bam*H I + *Eco*R V, R= *Bam*H I + *Pst* I. Ladder (GIBCO 1 kb ladder, Rockville MD). There are two different *Bam*H I digests per cosmid (L-R: multicore buffer – Promega, Madison WI and OPA buffer – Pharmacia, Piscataway NJ). Arrow indicates *Bam*H I + *Pst* I 4 kb digest of G22:A12 with the4 kb fragment containing *ncvip1*.

Figure 32b. G22:A11 and G22:A12 genomic cosmid digests hybridized to pCB1004
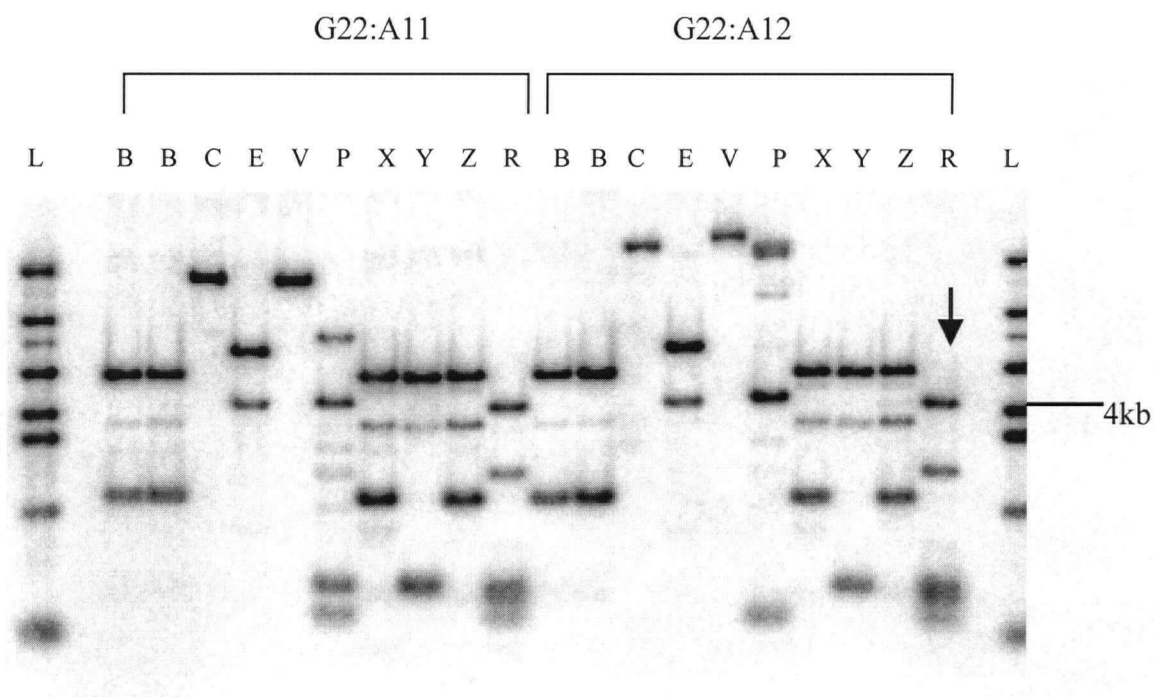
Figure 32b. G22:A11 and G22:A12 genomic cosmids from the Orbach/Sachs library probed with PCR amplified *ncvip1*. Blot exposed for 10 minutes on a Phosphorimager screen. L= ladder, B= *Bam*H I, C= *Cla* I, E= *Eco*R I, V= *Eco*R V, P= *Pst* I, X= *Bam*H I + *Cla* I, Y= *Bam*H I + *Eco*R I, Z= *Bam*H I + *Eco*R V, R= *Bam*H I + *Pst* I. Ladder (GIBCO 1 kb ladder, Rockville MD). There are two different *Bam*H I digests per cosmid (L-R: multicore buffer – Promega, Madison WI and OPA buffer – Pharmacia, Piscataway NJ). Arrow indicates *Bam*H I + *Pst* I 4 kb digest of G22:A12 with the 4 kb fragment with *ncvip1*.

A region of G22:A12 was sequenced using *ncvip1* cDNA-specific primers MDH1, MDH2, MDHS1 and MDHS2. A 330 base pair stretch aligned with the cDNA *ncvip1* sequence and identified a 65 bp intron. After 330 bp, there was sequence degeneracy and reverse sequencing did not match at all with the cDNA *ncvip1* sequence (data not shown). Attempts to

PCR amplify the *ncvip1* fragment from G22:12 using *ncvip1* cDNA-specific primers were unsuccessful, yielding fragments smaller than the expected size 1 kb fragment (data not shown).

From the sequencing and PCR data, it was determined that the G22:A12 cosmid contained a partial *ncvip1* sequence and was not subject to further characterization or subcloning efforts.

3a.3.2.1.2 Sequencing *ncvip1* from genomic DNA

The overall genomic sequence for *ncvip1* was verified by amplifying and sequencing *ncvip1* in strain 201, using MDHF and MDH2 primers; the 201 sequence matched the predetermined sequence from the *ncvip1* cDNA, with the added intron identified in the sequencing of G22:A12. The sequence of the *ncvip1* was also verified with comparison to contig 1.396 from the Whitehead Institute *N. crassa* genome sequencing project (website in **Figure 10**); TBLASTN of the Whitehead genome sequence with NCVIP1 yielded one hit, which was identical.

3a.3.2.1.3 Characterization of the *ncvip1* gene sequence

Similar size protein products from the cDNA and Vip1, an uncharacterized protein from *S. pombe* (E value 7e-31 in a BLASTP search), determined the 917 bp open reading frame (ORF) for *ncvip1* (**Figure 33**). The upstream region was not investigated until the release of the *N. crassa* genome sequence from the Whitehead Institute (website found in **Figure 10**).

Upon examination of the upstream region in genomic contig 1.396, it was found that there were no putative CAAT or TATA boxes within the expected ranges of –30 and –80 respectively (**Figure 33**; Bruchez *et al.*, 1993). All – and + reference numbers are in relation to the translational ATG start sequence because the transcriptional start site is unknown. Putative CAAT boxes are present at –845, -835, -680 and –543 bp and a TATA box at –799 bp relative to the ATG (**Figure 33**). A putative transcriptional start site (normally located at +1 bp) is found at –194 bp from the ATG; the sequence matches the consensus sequence except for one base, giving CCATCATC versus TCATCANC (Bruchez *et al.*, 1993). A translational start site consensus of CAMMATGGCT (where M is a C or an A – Edelmann and Staben, 1994), is semi-conserved, as demonstrated in the sequence of CAATCATGTCT, spanning the ATG. In the sequence in *ncvip1*, there is an inserted base (likely the T) and there is a T following the ATG in place of the G. Taking all of the upstream sequence into consideration, it is possible that the *ncvip1* gene is in fact larger than examined in this study, although visual inspection and a protein prediction program could not deduce an alternative ATG start codon. Once exception was a prediction using the BCM Gene Finder program (http://searchlauncher.bcm.tmc.edu:9331/gene-finder/gf.html), which predicted a polypeptide of 357 aa, but contained a 449 bp intron which is exceptionally large for *N. crassa*. For the purposes of this study, the 283 aa polypeptide will be the one characterized, as an analysis of protein motif search programs did not demonstrate any regions of significance.

Figure 33. *ncvip1* and NCVIP1 alignment

```
-858  GCTATGCGTAGACAATGGCATTCAAATCCCAAAGCACACCAAAGTGAAGACAGCAGGATCATATAATTAAGTACCTCTTACCTCTCCACTAGTAC
-755  GAACGAAACACCAAACTACTTGAAACGTCAACACACTGGACATGCTCAGCAAAGACTCAGCAGAAGACTTCAGTTTTGAGGCGGAATTTCA
-672  CATGCTACTTGTACAGCATTGTGATATTGCGATGTGCAGCTTCATCCCAGCGGAGATGAGGAGCCTACCTACTCAAACACCGTCGTCACTT
-579  CCATCCTGAGTCTACATAGGTACCTCTCTACACTACAATTTCCAGCGGACAGATGGGGCCTTTGCCCCTGCTAACCAACTTCCTGACGTGT
-486  CTTGGCGGTTGAAGCTTAGCTCCCCTCGGCCCTCTGGCCTGGCTGCCGTGTACCCCTCTTTCCGCCCATGGTTTCCGCCCTCGTTCGCACAGCC
-393  ACCCTGGATGGCCCCTGATCATGCGATGACGATCGCCTGGTCGCCGTGTACCCCTCTTTCTCCAGCTTTCTATCATCC
-300  CTCGCTCCAGCCTACTGTAGTGTAGCAGTGACAGTCAAGTCGAGTTGGTTGTGAACCAGTCAAGTCCACTGCCCTGTATTTTTTTACACTTT
-207  GGGCAAGGTAATCCAACCAACTGACGAACCCTGGTACATACATAAGACCCAGTTCCTTCCTTTGACTTCTGCTTTTGGCTCTAC
-114  TTAGAATCTCTCTCCATTGCCACAAGAAACCAAACATTTCGGTTCGACCACTAGCACAAACGCACACAGAT
 -21  ACCATCTCAACCACCAGCAATCATGTCTACAGTCTACGTCAAGATCGCCCAACACTGAGGAGAAGGACATTCGTGCCTCTTCAGCTTC            24
                       M   S   T  [V   Y   V   K   N   I   G   A   N   T   E   E   K   D   I   R   A   F   F   S   F
 +73  TGGTAAGCTTCTTGTGTCCCAAAAAGATGGATGACAGGGATGCAAGCTAACGATGAACCCGGAACACAGCGGCAAGATCAGCTCCCTCGATGTCA    33
      C                                                                       G   K   I   S   S   L   D   V
+166  CCACCGAAGGCGGAGACCAAGTCCGCCACGGTGACCTTTGAAAAGGAGGAGCGCCGCTCCTTCCTCTCGATCACACCAAGCTTGGCG            64
      T   T   E   G   E   T   K   S   A   T   V   T   F   E   K   E   S   A   A   R   T   A   L   L   L   D   H   T   K   L   G
+259  AGCACGAGCTCTCTGTCACCTCTGCTTCGGGAGAGCACGCCGATTCCGGGACAAATGTCCACCCAAGTCCGGACACAGATG            95
      E   H   E   L   S   V   I   T   S   A   S   G   E   H   A   D   S   G   D   N   V   H   P   K   S   D   A   D   R   D   T   D
+352  AGATCACACAGGAGGAGAAGCCCCGCGCGGTCTTCGAGTACTTGGCCAGCGGCTACCTCGTTGCCGACAGCGGCCTGAAGACGGCCA           126
      E   I   T   Q   E   E   K   P   R   A   R   V   L   A   E   Y   L   A   S   G   Y   L   V   A   D   S   G   L   K   T   A
+445  TTGCTCTCGACGAAGACACGGCGGCTCTCTCAGCGCTTCCTGTCGACAATCCAGAACCTGACCAAAAGTACCACGCACCGACCGGCCAAGA          157
      I   A   L   D   E   K   H   G   V   S   Q   R   F   L   S   T   I   Q   N   L   D   Q   K   Y   H   A   T   D   R   A   K
+538  CGGCCGACCAGAGCTACGGCATCACTGCGCGGGTCAACTCCCTGTTCCACGAGGCCAAGAAGCCTTGGAGGCGCCTGGCG           188
      T   A   D   Q   S   Y   G   I   T   A   R   A   N   S   L   F   S   G   L   S   S   Y   F   E   K   A   L   E   A   P   G
+631  CCAAGAAGATTGTCGACTTCTACACCACCGGCTCCAAGCAGGTGCAGGACATCCACAACGAGGCCAAGCGCCTGGCTGAGCTGAAGAAGCAGG          219
      A   K   K   I   V   D   F   Y   T   T   G   S   K   Q   V   Q   D   I   H   N   E   A   K   R   L   A   E   L   K   K   Q
+724  AGGCCGGCGGCTCGTCGTACAAGGCTGCGGCCAAGCTGGACAAGATCTTCGGCGGCCGAGAAGGGCCCCGCCCAGGAGAGCAAGCCCAATGACCAGG       250
      E   A   G   G   S   Y   K   A   A   G   L   D   K   I   F   G   A   E   K   A   P   G   Q   E   S   K   P   N   D   Q
+817  TGCCCGGCGCTGCCCCTTCGGATGGCTGCCGCCACCGAGTCTAACCAGCAACCGATTTCCGAGGAGCGATTTCCGGACTGCCGAGAAGATTC         281
      V   P   G   A   A   P   S   D   A   A   A   T   E   S   N   Q   Q   P   I   S   E   G   A   Y   P   G   T   A   E   K   I
+910  CCCAGTAAAGAAATGGTAGCGAGGCAGAATGGAGGTGGATCACTGGTTGACGACACGCGCTTCGCATGGTTAGCTGAGCCAGCAAGGTGTTTG          283
      P   Q

+1003 GCTTTTCGCTCGCTGATTTCATGATACTTTTATGACTGGTTATGGGACTGGGATTTCATTGTTGGCGTTGCCGTTGCCTCTAAATCTTTAAGTAATGAACA
+1096 AAATTCTCATCATTTGAAGCTTATCACGGCGGTAGTACCTTGCAGTAGCATAATGACGGTACATGCAATGTTCGATCAAGAAGCATGATCAAG
```

Figure 33. The upstream region of *ncvip1* does not conform to consensus translational start sites and the placements of putative CAAT and TATA boxes. All – and + reference numbers are in relation to the translational ATG start sequence because the transcriptional start site is unknown. CAAT boxes (highlighted and underlined; found at –845, –835, –680 and –543 bp respectively). A putative TATA box is found at –799 bp (highlighted and underlined). A consensus transcriptional start site is located at –194 bp (wavy underlined, bold and highlighted – Bruchez et al., 1993). From the transcriptional start codon (ATG, underlined), there is a putative consensus with a translational start site (outlined – Edelmann and Staben, 1994). One 65 bp intron is noted (italicized, with the 5' (+74 bp), lariat (+118 bp) and 3' (+135 bp) splice site and consensus sites in underlined italics – Bruchez et al., 1993). The termination cosensus (+913 bp = AGTAA; bold thick underlined and highlighted – Bruchez et al., 1993). The putative RNA-recognition motif is bracketed ("[ ]") in the NCVIP1 sequence (from SMART search with BLASTP of the NCBI GenBank database; http://www.ncbi.nlm.nih.gov/BLAST/). The polyadenylation signal sequence (+ 1088 bp = bold – Bruchez et al., 1993). DNA bp reference numbers are on the right; protein amino acid reference numbers are on the left.

N. crassa is known to have small introns (on the order of 60-80 base pairs), with larger introns being rare (Bruchez *et al.*, 1993). There is one concrete intron in *ncvip1*, spanning 65 bp (**Figure 33**), as established by comparing the cDNA versus genomic DNA sequences. The intron splicing 5' and 3' consensus sequences (+74 and +135 bp respectively, with reference to the ATG) match well with those published by Bruchez *et al.* (1993). The consensus for the intron 5' sequence is G-GT(A/G)(A/C)G(T/C) (Bruchez *et al.*, 1993); the 5' intron region in *ncvip1* is GGTAAGC, varying by only one base. The intron 3' region is a bit more variable from the consensus of G(A/T)(T/C)AGG (Bruchez *et al.*, 1993). In *ncvip1*, the intron 3' splice sequence is AACAGC. The first and last bases vary, but these bases are found in a number of other analyzed intron sequences (such as *cpi, crp-1, crp-3* and *vma-1* – Bruchez *et al.*, 1993), which match the 3' for *ncvip1* at all but the last base. The lariat consensus (+118 bp relative to the ATG) is (G/A)CT(A/G)AC (Bruchez *et al.*, 1993); the lariat sequence in *ncvip1* is exact (GCTAAC).
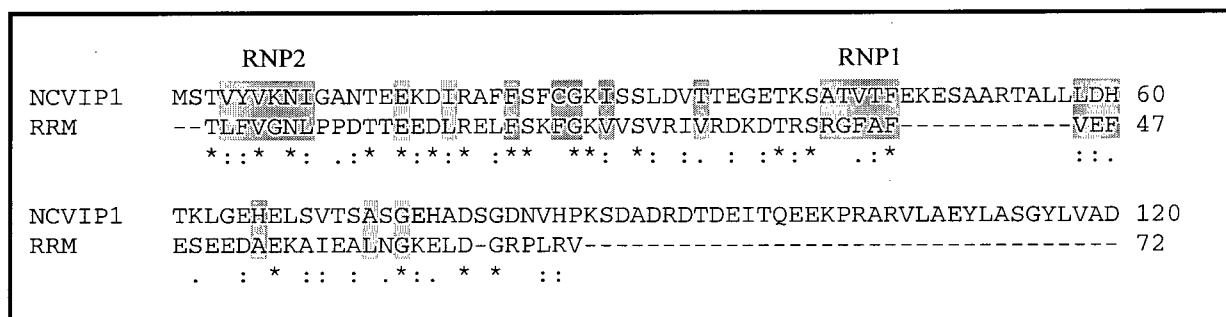
The last regulatory region is the 3' termination sequence, with a consensus of AATAAA (Bruchez *et al.*, 1993). The 3' termination sequence in *ncvip1* varies slightly, AGTAAA (+913 bp relative to the ATG; **Figure 33**), but matches that identified for *atp-2* (Bruchez *et al.*, 1993). A polyadenylation signal is found at position +1088 bp, with a sequence of AATGAA. The consensus for the 3' polyadenylation signal is AATAAA, but AATGAA is the 3' regulatory signal for *atp-2, ilv-2* and *nit-3* (Bruchez *et al.*, 1993).

3a.3.2.1.4 Characterization of the NCVIP1 polypeptide sequence

Translation of the 917 ORF of *ncvip1*, with the exclusion of the 65 bp intron, yields a putative polypeptide of 283 amino acids (**Figure 33**). A number of protein motif scan programs and analytical tools were used to analyze NCVIP1, but only those that gave informative results are included in the results. Programs used were: GCG (user name required); PSORT II; Blocks; Pfam; and SMART (via BLASTP with NCBI GenBank). All websites are listed in **Figure 10**.

In the Pfam, SMART and Blocks programs, an RNA recognition motif (RRM) was identified, with E values of 3.3e-10, 2.9e-06 and 1e-09 respectively. The RRM generally consists of a conserved region of 90 amino acids, with two RNP (ribonucleoprotein) sequences separated by 25-35 amino acids (Burd and Dreyfuss, 1994; Adam *et al.*, 1986). RRM motifs are conserved in a large number of species and RNA-binding proteins, which can associate with hnRNPs and snRNPs and have roles in splicing, localization and transport of RNA (Birney *et al.*, 1993). It is possible that a protein can have multiple RRM's, but NCVIP1 only contains one at the N-terminal region. LA proteins also contain an N terminal RRM; these proteins are the autoantigens involved in the autoimmune disease Systematic lupus erythematosus (SLE – Keene, 1996). The RRM 3D structure has been determined through crystallization and NMR and has a βαββαβ domain fold (Nagai *et al.* 1990; Hoffman *et al.*, 1991). Protein binding to RNA is likely due to charged interactions and structural conformation (reviewed in Kenan *et al.*, 1991; Burd and Dreyfuss, 1994). An alignment of NCVIP1 with the consensus for the RRM is found in **Figure 34**.

Figure 34. NCVIP1 alignment with a RRM motif .

```
                 RNP2                                    RNP1
NCVIP1    MSTVYVKNIGANTEEKDIRAFFSFCGKISSLDVTTEGETKSATVTFEKESAARTALLLDH  60
RRM       --TLFVGNLPPDTTEEDLRELFSKFGKVVSVRIVRDKDTRSRGFAF----------VEF  47
             *::*  *:  .:*  *:*:*  :**   **:  *:  :.  :  :*:*   .:*        ::.

NCVIP1    TKLGEHELSVTSASGEHADSGDNVHPKSDADRDTDEITQEEKPRARVLAEYLASGYLVAD 120
RRM       ESEEDAEKAIEALNGKELD-GRPLRV---------------------------------  72
              .   :  *  ::  :  .*:.  *  *    ::
```

Figure 34. ClustalW sequence alignment (http://www2.ebi.ac.uk/clustalw/) of NCVIP1 and an RRM consensus sequence as identified in the BLASTP search of the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/BLAST/) using the SMART program. "*" = identical or conserved residues in all sequences in the alignment; ":" = indicates conservative substitutions; "." = indicates semi-conservative substitutions. Highlighted regions denote the residues conserved in the RRM, with RNP1 and RNP2 being the two short sequences in the RRM in conjunction with the scattered mainly hydrophobic amino acid residues (described in detail in Burd and Dreyfuss, 1994)

The PSORT II prediction program for targeting signals and localization contains two prediction methods for localization: the Reinhardt's method for cytoplasmic/nuclear discrimination and the k-NN prediction method. Both methods predict NCVIP1 as a nuclear protein, with percentages of 89% and 60.9% respectively. The k-NN prediction method lists mitochondrial localization at 26.1%. These predictions correlate with the presence of the RRM motif, since RNAs are found in the nucleus.

A variety of other protein characteristics were examined concerning NCVIP1 – such as coiled-coil regions and secondary structure predictions, yielded no additional information about NCVIP1. The predicted molecular weight for NCVIP1 is 3.0322 KD and the predicted PI is 5.28 according to ProtParam (website listed in ExPasy – website in **Figure 10**; **Figure 35a**). A list of the amino acid composition as determined by ProtParam is found in **Figure 35a**. A secondary structure prediction, using nnpredict is found in **Figure 35b** (program is listed in ExPasy, website in **Figure 10**).

Figure 35a. Amino acid composition, MW and PI for NCVIP1

```
Amino acid composition:

Ala (A)   37      13.1%      Leu (L)   19     6.7%
Arg (R)    9       3.2%      Lys (K)   24     8.5%
Asn (N)    8       2.8%      Met (M)    1     0.4%
Asp (D)   19       6.7%      Phe (F)    9     3.2%
Cys (C)    1       0.4%      Pro (P)   10     3.5%
Gln (Q)   13       4.6%      Ser (S)   27     9.5%
Glu (E)   24       8.5%      Thr (T)   21     7.4%
Gly (G)   21       7.4%      Trp (W)    0     0.0%
His (H)    7       2.5%      Tyr (Y)    9     3.2%
Ile (I)   12       4.2%      Val (V)   12     4.2%


Asx (B)    0       0.0%      Glx (Z)    0     0.0%

- charged res.(Asp + Glu): 43
+ charged res.(Arg + Lys): 33
_____

MW = 3.0322 KD                Theoretical PI = 5.28
```

Figure 35a. NCVIP1 protein characterization according to ProtParam, a proteomic program in ExPasy (http://www.expasy.ch/prosite/). Included are the amino acid composition, molecular weight and predicted PI.

Figure 35b. Predicted secondary structure for NCVIP1

```
Secondary structure prediction (H = helix, E = strand, - = no prediction):
---EEEE-------HHHHHHHHHH--------E----------EEHHHHHHHHHHHHHH-H
--H---HEEE----------------------------------HHHHHHHHHH-EEEE-
---HHHHHHH-------EEHEHHHHH-H-H------------H-EEEE---EE----HHH
HHHHH-----EEEEEEH------EHHHHHHHHHHHHHHHH------E-H----HHH-----
-----------------------------------------------------
```

Figure 35b. Secondary structure prediction for NCVIP1 from nnpredict (program part of the ExPasy website – **Figure 10** for website) for the 283 amino acid polypeptide. Included are the α-helices (H), β-sheets (E) and non-predicted regions. (-)

A BLASTP NCBI GenBank database search (website in **Figure 10**) NCVIP1 yields one match with an E value equal to or less than e-10 in TBLASTN searches. This match is Vip1 from *S. pombe*, with an E value of 7 e-31 (also described in more detail in section **3.3.1**). The significance of this similarity remains unclear, as is the function.

NCVIP1 has similarity to a large number of sequences isolated from evening and morning expressed ESTs (expressed sequence tags isolated from a time course experiment), as found in BLAST searches (website in **Figure 10**; BLAST results in **Figure 36**). Only those with E values equal to or less than -05 were included. These genes are putatively expressed in response to circadian rhythm, so it may be that *ncvip1* is regulated with a circadian clock. However, there is no experimental evidence to support this claim.

Figure 36. TBLASTN of NCVIP1 against the *N. crassa* evening and morning expressed ESTs

```
Query=  Ncvip1(283 letters)


Sequences producing High-scoring Segment Pairs:     E value
e1h12ne.f1                                          5.6e-73
g4b11ne.f1                                          1.2e-69
e1b09ne.f1                                          3.4e-69
e1h11ne.f1                                          3.4e-69
e7a05ne.f1                                          3.4e-69
g4c05ne.f1                                          3.4e-69
g3a09ne.f1                                          3.4e-69
d3g05ne.f1                                          1.6e-68
e3c04ne.f1                                          5.8e-61
d1a08ne.f1                                          6.9e-61
g6d11ne.f1                                          8.1e-60
g3a09ne.r1                                          2.7e-38
h7f02ne.f1                                          1.9e-32
d1a04ne.r1                                          1.3e-26
e3c04ne.r1                                          1.3e-26
c9d12ne.f1                                          2.3e-06
```

Figure 36. TBLASTN (Altschul *et al.*, 1997) of NCVIP1 with the evening and morning expressed EST database for *N. crassa* (http://www.genome.ou.edu/fungal.html). EST numbers are listed with the probability score (E value). Only hits with E<-05 were included.

3a.3.2.2 RFLP mapping of *ncvip1*

To determine the genomic location of *ncvip1*, a method of restriction fragment length polymorphisms (RFLP) was employed. This method is feasible for mapping genes in *N. crassa* that have been cloned, in contrast to classical methods of mapping via percent crossovers and recombination of marker genes. For this method, a phenotype or linked marker that can be scored is required. The *Hin*d III blot used in this study contained the parental Mauriceville-1c and Oak Ridge strains and 38 progeny. The set of genomic DNA digests provides an M/O (Mauriceville/Oak Ridge) polymorphic RFLP pattern based on inheritance of polymorphisms within a genomic region compared to the parental strains (description in section **3a.2.2.5**).

The *Hin*d III blot was probed with a probe made from the genomic cosmid G22:A12, from the Orbach/Sachs pMCosX cosmid library. This cosmid was found to contain a partial sequence of *ncvip1* (section **3a.3.2.1.1**). The pattern generated can be seen in **Figure 37**; the interpretation reads: OMMMMMOMOOMMMMOOOOOOOO?O?MXO?OMMMMOOMM. This pattern was compared with the pattern of known genes in an RFLP mapping database located on the *Neurospora crassa* genome homepage (website in **Figure 10**), localizing *ncvip1* to left arm of linkage group VI, close to the centromere. The closest gene identified to *ncvip1* was *ccg-12*, a clock-controlled gene (website in **Figure 10**).

Figure 37. RFLP *Hin*d III mapping blot probed with G22:A12 containing *ncvip1*
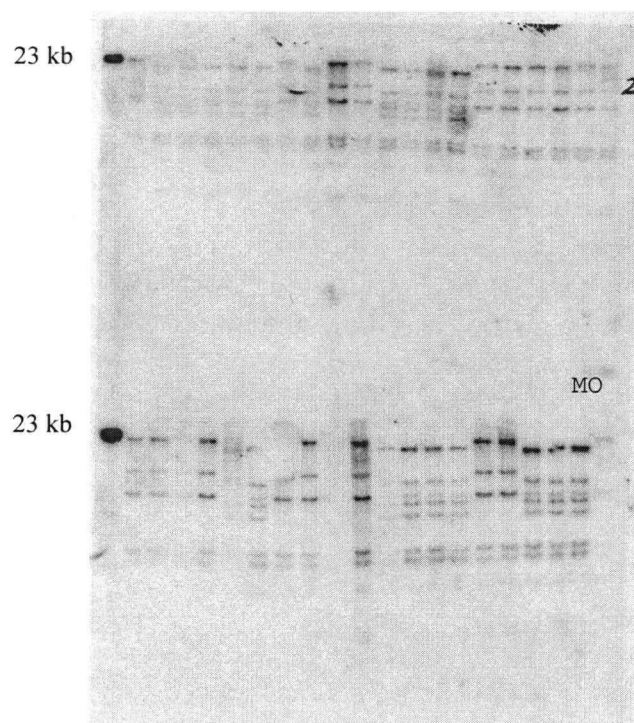
Figure 37. RFLP mapping blot (*Hin*d III digested genomic DNA samples. Blot probed with genomic cosmid G22:A12 and exposed for 8 days, using X-ray film. The lanes are: (top) marker (*Hin*d III/*Eco*R I λ DNA), A1, A4, B6, B7, C1, C4, D5, D7, E1, E3, E7, F1,F3, G1, G4, H5, H7, I6 and I8; (bottom) marker, J1, J4, K1, K4, L1, L4, M5, M8, N2, N3, O2, O4, P1, P4, Q2, Q4, R1, R4, Mauriceville (M) and Multicent (O). Parental digests are indicated on the blot (M and O). M/O pattern from top (L-R):
OMMMMMOMOOMMMMMOOOOOO
; bottom (L-R):
OO?O?MXO?OMMMMOOMM
The pattern does not include the last two lanes on the bottom, as they represent the parental RFLP patterns. "?" represents an indistinguishable pattern.

3a.3.2.3 Creating a *ncvip1* mutant and the corresponding analysis
3a.3.2.3.1 Initial *ncvip1* RIP cross results: "MH", "B", "C", "R", "S" and "T" transformation strains

It was decided to mutate *ncvip1* via **r**epeat **i**nduced **p**oint (RIP) mutation analysis because there was a higher mutation rate versus the alternative homologous recombination approach (10% homologous recombination rate in *N. crassa*). When project was initiated, the genomic clone was unidentified and the location in the genome was also unknown. This made it impossible to create a suitable construct for transformation and homologous recombination.

RIP analysis was used to create mutants though triplicate transformations of *ncvip1* into 201 (*his-5 trp-4 tol; pan-2 a*) and I-20-26 (*ad-3B arg-1 A*) and pCB1004 into 201 and I-20-26.

As described in section **3a.2.2.6.4**, crosses were made and progeny for the 201 transformations picked and screened (**Table 6**). Of note, all lab book and strain references in the Glass lab are currently labeled as "A" in place of "MH"; the change was made to avoid confusion with mating-type designations. Ascospore viability in all crosses appeared to decrease over time, limiting the number of progeny that could be isolated for analysis. This issue is addressed in section **3.4.4** in the discussion. Once crossed, four major classes of progeny are theoretically generated, assuming the RIP of the gene of interest creates a null mutation: 1) $tol^- + ncvip1^-$, 2) $tol^- + ncvip1^+$, 3) $tol^+ + ncvip1^-$, and 4) $tol^+ + ncvip1^+$. If *ncvip1* was an essential gene, classes 1) and 3) will not be represented, unless the degree of mutation in *ncvip1* results in a partial-loss -of-function mutant.

Table 6. *ncvip1* **RIP cross screening**

| CROSS*	# PICKED	# SCORED FOR NUTRITIONAL REQUIREMENTS AND HYGROMYCIN	# OF PROGENY WITH NUTRITIONAL REQUIREMENTS	# OF HYGROMYCIN SENSITIVE PROGENY	# SCREENED FOR RIP
"MH"	272	67	50	57	47
"B"	135	43	22	43	
"C"	272	60	47	17	16
"R"	167	7	4	7	
"S"	105	8	6	6	
"T"	167	15	4	8	

* Note: Details of strains used in crosses are: "MH" = 201 (*his-3 trp-4 tol; pan-2 a*)/*vip*1.2.12.2 x FGSC 5448 (*ad-3A; inl A*); "B" = 201/*vip*1.13.5 x FGSC 5448; "C" = 201/*vip*1.2.7.4 x FGSC 5448; "R" = 201/PCB6.2 x FGSC 5448; "S" = 201/PCB7.5 x FGSC 5448; "T" = 201/PCB 2.3 x FGSC 5448. Numbers after *vip*1 and PCB denote the corresponding construct.

3a.3.2.3.2 Screening for *ncvip1* RIP progeny: "MH" and "C" crosses

"MH" and "C" crosses ("MH" = 201 [*his-3 trp-4 tol; pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 [*ad-3A; inl A*]; "C" = 201/*vip*1.2.7.4 x FGSC 5448) were chosen for further examination due to

fewer ectopic integrations of the *ncvip1* plasmid, as assessed by Southern blot (data not shown). Since there was no associated phenotype on Vogel's minimal medium slants (Vogel, 1964) provided with the appropriate supplements, "R", "S" and "T" control pCB1004 transformation cross progeny were not subject to further analysis ("R" = 201/PCB6.2 x FGSC 5448; "S" = 201/PCB7.5 x FGSC 5448; "T" = 201/PCB 2.3 x FGSC 5448). These crosses were completed to provide a comparison for "MH", "B" and "C" cross progeny to support phenotype association with RIP of *ncvip1* and not from integration of the plasmid construct into the genomic DNA. It should also be noted that the original crosses established to generate RIPed *ncvip1* progeny did not contain pantothenic acid in the crossing medium (despite this requirement with the *pan-2* marker), so there was a bias against selection for *pan-2* progeny. Of interest to this study, MH27 (*pan-2 a*) had an initial poor growth that was relieved after serial subcultures. This may be significant in interpreting the appearance of the novel mutation resulting in partial suppression of mating-type associated vegetative incompatibility (sections in **3b.3**). Screening for RIP through digesting genomic DNA samples with *Dpn* II and *Sau*3A I was completed for 47 "MH" progeny and 16 "C" progeny – data not shown). A sample blot is given in **Figure 38** to demonstrate the controls used and the difference in restriction patterns.

Figure 38. RIP screening with genomic *N. crassa* DNA digested with *Dpn* II and *Sau*3A I
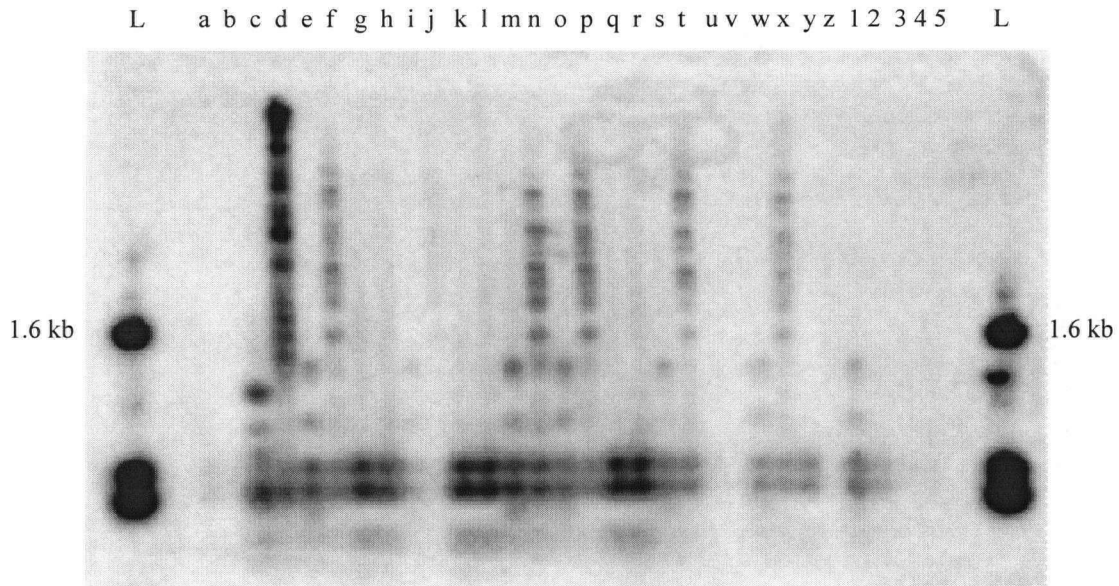
Figure 38. MH27 x FGSC 7214 (*pan-2 a x ad-8 lys-5 A*) progeny blot #1, probed with PCR amplified *ncvip1* and exposed overnight (Phosphorimager). Comb #1 - ladder (λ/*Hin*d III DNA ladder = L), blank, blank, a = 201 (D), b = 201 (S), c = "MH" (D), d = "MH" (S), e = **MH27** (D), f = **MH27** (S), g = FGSC 7214 (D), h = FGSC 7214 (S), i = MH27.10 (D), j = MH27.10 (S), k = MH27.11 (D), l = MH27.11 (S), m = **MH27.14** (D), n = **MH27.14** (S), o = **MH27.19** (D), p = **MH27.19** (S), q = MH27.23 (D), r = MH27.23 (S), t = **MH27.26** (D), u = **MH27.26** (S), v = MH27.27 (D), w = MH27.27 (S), x = **MH27.28** (D), y = **MH27.28** (S), z = MH27.45 (D), 1 = MH27.35 (S), 2 = MH27.35 (D), 3 = MH27.45 (S), 4 = MH27.55 (D), 5 = MH27.55 (S), blank, blank, ladder. D= *Dpn* II; S= *Sau*3A I. See section **3.2.2.6.13** for methods and conditions. Strains with methylation changes are bolded in the text.

None of the 16 "C" ("C" = 201 [*his-3 trp-4 tol; pan-2 a*]/*vip*1.2.7.4 x FGSC 5448 [*ad-3A; inl A*]) progeny screened had methylation changes or hybridized to the pCB1004 vector (data not shown). This may suggest that "C" had very few or no ectopic copies of the *ncvip1* plasmid remaining in the strain at the time of the cross, although the strain was capable of growing on hygromycin medium.

For the "MH" cross ("MH" = 201 [*his-3 trp-4 tol; pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 [*ad-3A; inl A*]), a total of 17 of the 47 progeny were identified as having alterations in the methylation patterns when probed against *ncvip1* PCR generated sequence (MH26, MH27, MH38, MH'12, MH'16, MH'17, MH'36, MH'42, MH'45, MH'48, MH'52, MH'59, MH86, MH108, MH118, MH139 and MH146 – data not shown). The evidence of methylation changes associated with RIP was complicated by the fact that the parental "MH" transformed strain demonstrated *de novo* alterations after the initial transformation, before the sexual cross. This was also observed in transformations involving *mat A-2* and *mat A-3* (L. Glass, personal communications). This may mean that the methylation pattern differences in progeny digested with *Dpn* II and *Sau*3A I may be the result of inheritance of the "MH" parental copy or from changes associated with RIP. If this is true, more than half of the "MH" progeny should have methylation changes, due to 1:1 segregation of the "MH" gene copy plus additional *ncvip1* RIPed strains. The number of 17/47 is much lower than expected. The reason for this misrepresentation is unclear, but may suggest that *ncvip1* RIPed progeny were not represented in the viable progeny pool that was subject to Southern analysis.

The 17 "MH" progeny also hybridized to the vector pCB1004, despite the selection against hygromycin resistant progeny, which maintained the vector (data not shown). This may suggest that the ectopic integration of the *ncvip1* transformation plasmid is linked to the resident copy of *ncvip1*.

3a.3.2.3.3 Genotypes, phenotypes and het test results of "MH" *ncvip1* RIP cross progeny

Genotypes of 16 "MH" progeny (201 [*his-5 trp-4 tol; pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 *ad-3A; inl A*) with methylation changes in *ncvip1* (minus MH'59) were assessed. The *trp-4* marker in the 201 transformed strains is closely linked to *tol* (1 mu), so *trp-4* requirements were used to assess *tol*. Genotyping and het test results of the 16 "MH" progeny can be seen in **Table 7**, with FGSC 7214 (*ad-8 lys-5 A*) or FGSC 5448 (*ad-3A; inl A*) as *mat A* testers and R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*) as the *mat a* tester. MH26 (*his-5 trp-4 tol A*), MH27 (*pan-2 a*), MH38 (*his-5; pan-2 A*), MH'12 (*pan-2; inl A*), MH'48 (*his-5; pan-2 A*) and MH86 (*his-5 trp-4 tol; inl; pan-2 a*) were taken through a cross with FGSC 7214 (*ad-8 lys-5 A*) for crosses with *mat a* strains and FGSC 4 (*lys-5 ylo-1 a*) with *mat A* strains. These crosses were completed to determine if the progeny were female fertile, and all were female fertile and produced ascospores. Viability of the ascospores was not assessed.

Table 7. "MH"* progeny potentially RIPed in *ncvip1*

"MH" PROGENY	GENOTYPE	HET TEST
MH26	*his-5 trp-4 tol A*	Incompatible with *mat a*
MH27	*pan-2 a*	Partial suppression of mating-type associated vegetative incompatibility
MH38	*his-5; pan-2 A*	Incompatible with *mat a*
MH'12	*pan-2; inl A*	Incompatible with *mat a*
MH'16	*his-5 trp-4 tol; pan-2 A*	Incompatible with *mat a*
MH'17	*pan-2 A*	Incompatible with *mat a*
MH'36	*his-5 trp-4 tol; pan-2 A*	Incompatible with *mat a*
MH'42	*his-5; inl; pan-2 A*	Incompatible with *mat a*
MH'45	*ad-3A; his-5 trp-4 tol; inl; pan-2 A*	Incompatible with *mat a*
MH'48	*his-5; pan-2 A*	Incompatible with *mat a*
MH'52	*his-5 trp-4 tol; pan-2 a*	Incompatible with *mat A*
MH86	*his-5 trp-4 tol; inl; pan-2 a*	Incompatible with *mat A*
MH108	*trp-4 tol; pan-2 a*	Inconsistent; likely incompatible
MH118	*his-5; inl; pan-2 a*	Incompatible with *mat A*
MH139	*his-5 trp-4 tol a*	Incompatible with *mat A*
MH146	*his-5 trp-4 tol; pan-2 a*	Incompatible with *mat A*

* "MH" = 201 [*his-5 trp-4 tol; pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 *ad-3A; inl A*

The het tests revealed a partial suppression of mating-type associated vegetative incompatibility with the strain MH27 (*pan-2 a*; **Figure 39**). The phenotype is called the **sup**pression of **m**ating-type **v**egetative **i**ncompatibility (Sup [mvi]) phenotype.

Figure 39. MH27 influence on mating-type associated vegetative incompatibility
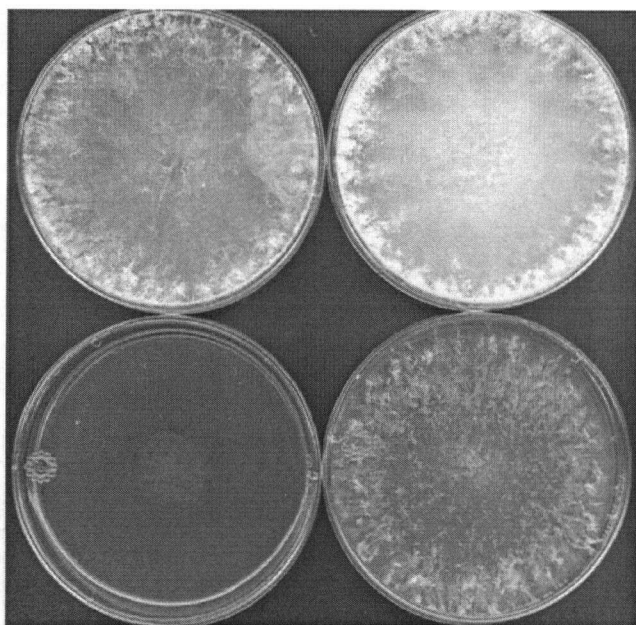
Figure 39. MH27 Sup (mvi) phenotype = partial suppression of mating-type associated vegetative incompatibility. Starting from top left and going clockwise: R1.21 + 201 (*pyr-4 arg-5 pregc-2 trp-3 a + his-5 trp-4 tol; pan-2 a*), MH27 + R1.21 (*pan-2 a + pyr-4 arg-5 pregc-2 trp-3 a*), MH27 + FGSC 7214 (*pan-2 a + ad-8 lys-5 A*) and FGSC 7214 + R1.21 (*ad-8 lys-5 A + pyr-4 arg-5 pregc-2 trp-3 a*). Conidial suspensions of strains were co-inoculated onto Vogel's minimal medium (Vogel, 1964).

3a.3.2.3.4 Growth rates for individual strains: wild type, transformed parentals and "MH" cross progeny

Growth rates for wild type strains 201 (*his-5 trp-4 tol; pan-2 a*), R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*) and FGSC 5448 (*ad-3A; inl A*) were determined at room temperature (24°C) and 10°C (**Table 8**). These rates were compared with the 201 transformation strains ("MH" and "C" transformed with *ncvip1* and "R" transformed with pCB1004 - "MH" = 201 [*his-3 trp-4 tol; pan-2 a*]/*vip*1.2.12.2 x FGSC 5448 [*ad-3A; inl A*]; "C" = 201/*vip*1.2.7.4 x FGSC 5448; "R" = 201/PCB6.2 x FGSC 5448) and to 7 of the 17 progeny from the "MH" cross with methylation changes. All trials were completed in triplicate, with the exception of R1.21, which had four trials.

Table 8. Growth rates for controls, transformed parentals and MH strains at room temperature and 10°C

STRAIN	GROWTH RATE (cm/day) ± ST DEV (24°)*	GROWTH RATE (cm/day) ± ST DEV AT 10 °C*
201	6.57 ± 0.23	2.43 ± 0.11
R1.21	7.08 ± 0.72	ND
FGSC 5448	7.49 ± 0.13	2.51 ± 0.16
"MH"	6.91 ± 0.21	2.62 ± 0.33
"C"	7.13 ± 0.23	2.32 ± 0.33
"R"	6.70 ± 0.19	1.98 ± 0.67
MH26	7.09 ± 0.29	2.40 ± 0.13
MH27	6.24 ± 0.48	1.74 ± 0.54
MH38	7.02 ± 0.27	2.08 ± 0.26
MH'12	7.42 ± 0.17	2.11 ± 0.12
MH'16	7.07 ± 0.30	2.31 ± 0.15
MH'17	7.68 ± 0.24	1.93 ± 0.57
MH'36	8.02 ± 0.14	ND

* Note: days measured = 4, 5, 6. One exception = 201 (10° C), measured over days 8, 9, 10

All strains tested appeared to have a normal growth rate, as compared to the wild type controls (**Table 8**). The MH27 may have a slight growth defect, as noted in the slowest growth rates at both room temperature and 10°C. However, the variability at both temperatures was high, making the potential growth defect statistically insignificant.

3a.3.2.3.5 MH27 cross to segregate methylation of *ncvip1* from pCB1004

Because of the potential linkage of the ectopic integration of the transformation plasmid with *ncvip1*, a second cross was completed between MH27 (*pan-2 a*) and FGSC 7214 (*pan-2 a* x *ad-8 lys-5 A*). The same issues with ascospore viability that were encountered with the "MH" cross were evident, with viability of ascospores decreasing over time (section **3.4.4**). One hundred sixty-five progeny were heat shocked and picked (with the aid of Q. Xiang) to appropriate Vogel's slants (Vogel, 1964), yielding a total of 109 viable progeny. Genotyping

was completed for nutritional markers of these 109 progeny, but hygromycin sensitivity could not be used as a selection method as the original MH27 strain was hygromycin sensitive. Of the MH27 cross progeny isolated, 43 were screened for RIP. Again, blots were hybridized to PCR generated *ncvip1* and pCB1004 (data not shown; sample blot in **Figure 38**).

Twenty-four of 43 progeny had methylation changes (MH27.7, MH27.10, MH27.12, MH27.13, MH27.14, MH27.19, MH27.26, MH27.28, MH27.31, MH27.34, MH27.42, MH27.44, MH27.48, MH27.67, MH27.69, MH27.71, MH27.75, MH27.76, MH27.77, MH27.84, MH27.85, MH27.121, MH27.140 and MH27.199; **Table 9**), but unfortunately only one progeny with methylation changes and no vector sequence was isolated (MH27.69 – *pan-2 A*). This further supports that the ectopic integration is linked to the resident copy of *ncvip1*, resulting in a low frequency of recombination (1/24 = 4-5 mu from *ncvip1*).

As mentioned above, 109 progeny were generated from the MH27 (*pan-2 a*) x FGSC 7214 (*ad-8 lys-5 A*) cross for subsequent analysis, all of which were assessed for genotype and mating-type using the previously mentioned mating-type tester strains (*fl a* = *mat a* tester and FGSC 7214 + a^{m1} = *mat A* tester). Of these, only those also assessed in mating-type het tests are included in **Table 9** (59 strains in total). Het tests were completed as before, using R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*) and FGSC 5448 (*ad-3A; inl A*) as testers (**Table 9**).

Table 9. MH27 x FGSC 7214: genotypes, phenotypes and methylation analysis results

STRAIN	GENOTYPE	Sup (mvi) PHENOTYPE	VECTOR PRESENT	METHYLATION CHANGES
MH27.6	*pan-2 a*	yes	no	no
MH27.7	*pan-2 A*	N/A	yes	yes
MH27.8	*pan-2 A*	no	N/A	N/A
MH27.10	*ad-8 pan-2 a*	yes	yes	yes
MH27.11*	*ad-8 lys-5 a*	yes	no	no
MH27.12	*ad-8 pan-2 a*	yes	yes	yes
MH27.13	*pan-2 A*	no	yes	yes

Cont.				
MH27.14*	*ad-8 lys-5 a*	yes (but weak)	yes	yes
MH27.19	*pan-2 A*	no	yes	yes
MH27.23	*pan-2 A*	no	no	no
MH27.26	*ad-8 pan-2 A*	no	yes	yes
MH27.27	*ad-8 pan-2 A*	no	N/A	N/A
MH27.28	*ad-8 lys-5(?) A*	no	yes	yes
MH27.30	*pan-2 A*	no	no	no
MH27.31	*pan-2 A*	N/A	yes	yes
MH27.33	*lys-5 A*	no	N/A	N/A
MH27.34	*pan-2 A*	no	yes	yes
MH27.42	*pan-2 A*	N/A	yes	yes
MH27.44	*pan-2 A*	N/A	yes	yes
MH27.45	*pan-2 a*	yes	no	no
MH27.48	*pan-2 A*	no	yes	yes
MH27.55	*pan-2 A*	no	no	no
MH27.64	*lys-5 A*	no	no	no
MH27.66	*pan-2 A*	N/A	no	no
MH27.67	*wild type A*	no	yes	yes
MH27.69*	*pan-2 A*	N/A	no	yes
MH27.70	*pan-2 a*	yes	N/A	N/A
MH27.71	*pan-2 a*	yes	yes	yes
MH27.72	*ad-8 lys-5 a*	yes	N/A	N/A
MH27.73	*pan-2 A*	no	N/A	N/A
MH27. 75	*pan-2 a*	yes	yes	yes
MH27.76	*pan-2 a*	yes	yes	yes
MH27.77	*ad-8 lys-5 A*	no	yes	yes
MH27.78	*pan-2 A*	no	N/A	N/A
MH27.79	*pan-2 a*	yes	no	no
MH27.80	*pan-2 A*	no	N/A	N/A
MH27.81	*wild type A*	no	N/A	N/A
MH27.82	*pan-2 a*	yes	no	no
MH27.83	*ad-8 lys-5 A*	no	no	no
MH27.84	*pan-2 A*	no	yes	yes
MH27.85	*ad-8 lys-5 a*	yes	yes	yes
MH27.121	*ad-8 lys-5 a*	yes	yes	yes
MH27.124*	*lys-5 a*	no	N/A	N/A
MH27.125	*pan-2 a*	yes	N/A	N/A
MH27.126	*pan-2 a*	yes	N/A	N/A
MH27.127	*pan-2 a*	yes	N/A	N/A
MH27.131*	*ad-8 lys-5 A*	no	no	no
MH27.139	*pan-2 A*	no	no	no
MH27.140	*pan-2 A*	no	yes	yes
MH27.147	*ad-8 pan-2 a*	N/A	no	no

Cont.				
MH27.158	*pan-2 A*	no	N/A	N/A
MH27.169	*pan-2 a*	yes	N/A	N/A
MH27.173	*ad-8 lys-5 pan-2 a*	yes	N/A	N/A
MH27.175	*pan-2 a*	yes	N/A	N/A
MH27.189	*ad-8 pan-2 A*	no	no	no
MH27.191	*ad-8 lys-5 A*	no	no	no
MH27.192	*lys-5 a*	yes	no	no
MH27.193	*ad-8 lys-5 A*	no	N/A	N/A
MH27.199	*pan-2 A*	no	yes	yes

*Strains used in further analyses.

With the genotyping of MH27 progeny in **Table 9**, marker segregation was determined. In general, there appears to be an unequal segregation of markers (42/59 *pan-2*; 18/59 *ad-8*; 17/59 *lys-5*; 24/59 *mat a*; 35/59 *mat A*). The reasons behind this are unclear, but it may be a reflection of a bias in the progeny selected for genotyping or an influence of the Sup (mvi) phenotype on ascospore viability. There is only weak experimental evidence to support this argument, since the ascospore viability issues were also evident in the control crosses (section **3.4.4**). From the progeny genotyping, it is clear that the Sup (mvi) phenotype is independent of nutritional markers, as progeny with different marker combinations were isolated (i.e. MH27.11 [*ad-8 lys-5 a*], MH27.75 [*pan-2 a*] and MH27.173 [*ad-8 lys-5; pan-2 a*] - **Table 9**).

3a.3.2.3.6 MH27 Sup (mvi) phenotype relative to *ncvip1*

To further assess the partial suppression phenotype in MH27 (*pan-2 a*), *ncvip1* was sequenced using primers MDHF1 and MDH2. Surprisingly, there were no changes in *ncvip1* in MH27, indicating a novel mutation in this strain. *ncvip1* mutations may be recessive, which may explain why *ncvip1* potential RIP were not discovered in the initial het tests which would reflect

a heteroallelic condition. The alternative is that *ncvip1* does not influence mating-type associated vegetative incompatibility.

With the completion of the MH27 cross, it is likely that some of the progeny contain legitimate RIP mutations in the resident copy of *ncvip1*, although sequence analysis is required to confirm true mutants. Strain MH27.69 (*pan-2 A*), which has methylation changes observed in the hybridization to the *ncvip1* probe but not pCB1004, makes further investigation of *ncvip1* mutants possible. This strain does not have the dominant Sup (mvi) phenotype (**Table 9**). If MH27.69 is crossed to wild type strains, it should be possible to isolate progeny of both *mat a* and *mat A* genotype with different markers that are also *ncvip1* mutants, to assess the influence of *ncvip1* in a recessive situation in screening for affects on mating-type associated vegetative incompatibility. Due to time constraints, the creation of the MH27.69 strain is the completion of this study's investigation regarding the role of *ncvip1* in *N. crassa*.

3a.3.2.3.7 Overall summary of *ncvip1*

ncvip1 is a gene on linkage group VI encoding a putative protein product of 283 amino acids. RIP mutation analysis has generated one putative strain, MH27.69, with methylation changes (associated with RIP) and no hybridization to the vector, pCB1004. This strain will be available for future sequence analysis to determine if *ncvip1* is mutated. Once this has been established, MH27.69 will be crossed to generate additional progeny with different nutritional markers and *mat A* and *mat a* mating-types. The MH27.69 derived progeny will provide the tools to examine the influence of *ncvip1* on mating-type associated vegetative incompatibility.

3b.3.1 Sup (mvi) phenotype characterization

With the observation of partial suppression of mating-type associated vegetative incompatibility (Sup [mvi] phenotype) in MH27 (*pan-2 a*), and sequence analysis determining that *ncvip1* was not mutated, it was likely that MH27 contained a novel mutation that suppressed mating-type associated vegetative incompatibility. The MH27 strain was thus subjected to further analysis. Consideration of double mutants with *ncvip1* must be made in examining all experimental results with MH27 progeny. MH27 retained the transformation vector, as evidenced in the Southern hybridization with pCB1004 (data not shown), providing the means for RIP of *ncvip1* in any crosses completed with MH27.

3b.3.1.1 Localizing the mutation causing the Sup (mvi) phenotype

The main observation in the genotype and phenotype analysis was the segregation of the Sup (mvi) phenotype with MH27 *mat a* progeny (progeny genotyping in **Table 9**). All but one of the MH27 *mat a* progeny (MH27.124) had the Sup (mvi) phenotype, which was not evident in any of the MH27 *mat A* progeny. This linkage is very fortunate because it locates the mutation causing the partial suppression of mating-type associated vegetative incompatibility to *mat a*, either in *mat a* itself or a mutation that is tightly linked to *mat a*.

3b.3.1.2 Growth rates for controls and MH27 derived strains

Growth rates for MH27.11, MH27.14 and MH27.131 were determined at room temperature (24°C - **Table 10**). All strains tested appeared to have a normal growth rate,

134

compared with the control strains. All trials were completed in triplicate, with the exception of R1.21, MH27.11, MH27.14, and MH27.131, which had four trials each. As noted for the MH27 strain, there is a possible growth rate defect in MH27.14, but it is not statistically significant.

Table 10. Growth rates for controls and MH27 derived strains at room temperature

STRAIN	GROWTH RATE (cm/day) ± ST DEV (24°)*
201	6.57 ± 0.23
R1.21	7.08 ± 0.72
FGSC 5448	7.49 ± 0.13
MH27	6.24 ± 0.48
MH27.11	7.27 ± 0.79
MH27.14	6.84 ± 0.77
MH27.131	8.50 ± 0.76

* Note: days measured = 4, 5, 6

According to the growth rate results, all MH27 derived strains appear to grow normally at room temperature, with an average growth rate of approximately 7.14 cm/day (average includes the control growth rates).

3b.3.1.3 Heterokaryon test growth rates with MH27, MH27.11, MH27.14 and MH27.131

To further assess the degree of suppression of mating-type vegetative incompatibility associated with the Sup (mvi) phenotype, het tests were completed in race tubes (Ryan *et al.*, 1943), using R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*) and FGSC 5448 (*ad-3A; inl A*) as testers (**Table 11**). The MH27 progeny strains chosen for het tests did not contain the *pan-2* marker because of its leaky phenotype associated with *pan-2* when used as a het test forcing marker. The strains tested included MH27 (the original mutant; *pan-2 a*), MH27.11 (*ad-8 lys-5 a*), MH27.14 (*ad-8 lys-5 a*) and MH27.131 (*ad-8 lys-5 A*). The original parental MH27 data is

135

included, but must be viewed with caution because of a slowing in growth rate after four to five days at room temperature and large variability between trials (graphs not shown due to large variability between trials – variability represented through standard deviations in **Table 11**). It is unlikely that the growth arrest with the MH27 heterokaryon at room temperature is due to the *pan-2* marker, as the same slowdown in growth with large variability between trials was evident in the MH27.11 + FGSC 5448 het test (genotypes above – **Table 11**). As mentioned above, a graph of the slowed growth rate is not included because of the high variability reflected in the standard deviation. All heterokaryon trials were completed in triplicate, except those indicated without a standard deviation and the 10° C trials of MH27.14 + R1.21, MH27.14 + FGSC 5448, MH27.131 + R1.21 and MH27.131 + FGSC 5448, which had four trials each.

Table 11. Het test growth rates for controls, MH27 and MH27 derived progeny at room temperature and 10°C

HET TEST	GROWTH RATE (cm/day) ± ST DEV (24°C)*	GROWTH RATE (cm/day) ± ST DEV (10°C)*
FGSC 5448 + FGSC 7214 (*ad-3A; inl A + ad-8 lys-5 A*)	7.61 ± 0.49	2.83 ± 0.30
R1.21 + I-1-51 (*pyr-4 arg-5 pregc-2 trp-3 a + ad-3A nic-2 a*)	7.74 ± 0.51	2.23 ± 0.29
FGSC 5448 + R1.21 (*ad-8 lys-5 A + pyr-4 arg-5 pregc-2 trp-3 a*)	1.80	0.32 ± 0.30
MH27 + R1.21 (*pan-2 a + pyr-4 arg-5 pregc-2 trp-3 a*)	7.28 ± 0.46	3.12 ± 1.23
MH27 + FGSC 5448 (*pan-2 a + ad-3A; inl A*)	5.78 ± 1.12	3.28 ± 1.44
MH27.11 + R1.21 (*ad-8 lys-5 a + pyr-4 arg-5 pregc-2 trp-3 a*	8.48	ND
MH27.11 + FGSC 5448 (*ad-8 lys-5 a + ad-3A; inl A*)	6.64 ± 2.33	ND
MH27.14 + R1.21 (*ad-8 lys-5 a + pyr-4 arg-5 pregc-2 trp-3 a*)	7.77 ± 0.39	1.17 ± 0.53
MH27.14 + FGSC 5448 (*ad-8 lys-5 a + ad-3A; inl A*)	7.52 ± 0.82	0.36 ± 0.29
MH27.131 + R1.21 (*ad-8 lys-5 A + pyr-4 arg-5 pregc-2 trp-3 a*)	2.16 ± 0.19	0.19 ± 0.11
MH27.131 + FGSC 5448 (*ad-8 lys-5 A + ad-3A; inl A*)	8.04 ± 0.13	2.29 ± 0.15

* Note: days measured at 24°C = 4, 5, 6; 10° C = 3, 4, 5

As predicted, incompatible het tests reflected a slow growth rate, as demonstrated in the FGSC 5448 + R1.21 incompatible control and MH27.131 + R1.21 trials (**Table 11**). All other trials showed either a compatible growth rate (compared to the FGSC 5448 + FGSC 7214 and R1.21 + I-1-51 controls) or an intermediate rate with a high variability. It is unclear why there is a high variability in some test conditions, but it is noted only in the trials that have the Sup (mvi) phenotype, but an intact *ncvip1* (either no methylation changes [MH27.11] or verified with

sequence analysis [MH27]; **Table 11**). MH27.14 has the Sup (mvi) phenotype and methylation changes in *ncvip1* as well as vector sequence, so it would be interesting to determine if it is in fact a double mutant of *ncvip1* and the Sup (mvi) phenotype. It may be that the suppression of mating-type vegetative incompatibility is a result of the interaction between the gene products of the two loci.

Another interesting point in the growth rates of these heterokaryons is the rate observed at 10° C (**Table 11**). MH27 maintains partial suppression of mating-type associated vegetative incompatibility at this temperature and the growth rate does not seem to slow down after four days, as is apparent at room temperature (data not shown). MH27.14, a progeny of MH27 with different nutritional requirements and a potential double mutant with *ncvip1*, loses the partial suppression of mating-type associated vegetative incompatibility at 10° C that is observed at 24° C. The reasons behind this are unclear and more progeny (such as MH27.11) should be examined before definitive statements are made; the loss of suppression may in fact be a reflection of a double mutant phenotype. The reduced temperature may alter the interaction between gene products by stabilizing a complex that triggers incompatibility, which supports the model that protein complexes trigger vegetative incompatibility.

To determine if the heterokaryons are in fact true heterokaryons, hyphal tips of the heterokaryons from het tests with MH27, MH27.11, MH27.14 and MH27.131 were tested for mating-type. All heterokaryons with MH27 derived strain mated as *mat a* (data not shown), indicating a possible nuclear distortion ratio in the resulting heterokaryon. The influence of this observation on the Sup (mvi) phenotype is unclear. It is unlikely that the phenotype is the result of a cross-feeding reaction because there is no associated influence of the Sup (mvi) phenotype on *het-c* vegetative incompatibility. MH27, MH27.11, MH27.14, MH27.69 and MH27.173 (all

het-c Oak Ridge) were tested against *het-c* Panama strains and there was no suppression of *het-c* vegetative incompatibility.

The heterokaryon test results summary with MH27 derived strains appears to expose a possible nuclear distortion ratio defect in strains with the Sup (mvi) phenotype. Partial suppression of mating-type associated vegetative incompatibility appears to be more variable in strains defective only in gene causing the Sup (mvi) phenotype, but there may be an increased ability for suppression in a double mutant with *ncvip1*. The effect of temperature on the Sup (mvi) phenotype is unclear. Further assessment of the nature of *ncvip1* and the gene causing the Sup (mvi) phenotype in MH27 derived strains is necessary and more data is needed to make solid conclusions about the cause of the suppression of mating-type associated vegetative incompatibility.

3b.3.1.4 Cell death assays

Vegetative incompatibility is characterized by an increase in cell death (approximately 15-30% in incompatible heterokaryons or partial diploids with *het-6* and *het-c* – Jacobson *et al.*, 1998). Evan's Blue staining differentiates live and dead cells (Gaff and Okong'O-Ogola, 1971). Cell death assays were completed using MH27 and MH27 derived strains to determine the percentage of cell death in individual strains and heterokaryons. R1.21, I-1-51, FGSC 5448, MH27, MH27.11, MH27.14 and MH27.131 were assessed, with the same combination of het tests as examined in the growth rate analysis (section **3b.2.1.2**; results in **Table 12**). These het tests assessed the mutation causing the Sup (mvi) phenotype in strains containing different nutritional markers (MH27 versus MH27.14) and with a potential additional mutation due to RIP

in *nvip1* (MH27.14 versus MH27.11). Comparisons were made to wild type compatible (R1.21 + I-1-51 and FGSC 5448 + MH27.131) and incompatible controls (FGSC 5448 + R1.21 and R1.21 + MH27.131). MH27.131 is a progeny derived from the cross with MH27 and FGSC 7214 (*ad-8 lys-5 A*), with the same markers as MH27.14, but not containing the MH27 mutation or alteration in methylation in *ncvip1*. Genotypes for all of the strains used in the cell death assay are listed in **Table 12**. For this study, one time point was assessed and the cell death determined by counting 200 hyphal compartments and noting the number of cells containing stain. From these counts, the percentage of cell death was calculated.

Table 12. Cell death counts for individual strains and het tests

STRAIN OR HET TEST	CELL DEATH (# DEAD CELLS / # TOTAL CELLS X 100%)* = %	COMMENTS
R1.21 (*pyr-4 arg-5 pregc-2 trp-3 a*)	1/212 = 0.5%	
FGSC 5448 (*ad-3A; inl A*)	0/202 = 0%	
I-1-51 (*ad-3A nic-2 a*)	7/211 = 3%	
MH27 (*pan-2 a*)	2/202 = 1%	
MH27.11 (*ad-8 lys-5 a*)	7/211 = 3%	
MH27.14 (*ad-8 lys-5 a*)	0/209 = 0%	
MH27.131 (*ad-8 lys-5 A*)	6/215 = 3%	
R1.21 + FGSC 5448	16/216 = 7%	Incompatible control
R1.21 + I-1-51	3/216 = 1%	Compatible control
R1.21 + MH27	6/213 = 3%	Compatible
FGSC 5448 + MH27	49/215 = 23%	Partial suppression
R1.21 + MH27.11	4/204 = 2%	Compatible
FGSC 5448 + MH27.11	30/254 = 12%	Partial suppression
R1.21 + MH27.14	0/218 = 0%	Compatible
FGSC 5448 + MH27.14	22/233 = 9%	Partial suppression
R1.21 + MH27.131	12/236 = 5%	Incompatible control
FGSC 5448 + MH27.131	3/203 = 1%	Compatible

* Note: Numbers may be slightly higher than normal due to staining from manipulation of hyphae in staining process.

Individual strains and mating-type compatible het tests showed a low percentage of cell death, as expected ranging from 0-3% (**Table 12**). The figures above 1% may be inaccurate due

to damage from the staining process or because of the low number of cells and trials examined. Typically, there is very little to no cell death detected in individual strains or compatible heterokaryons (Q. Xiang and S. Sarkar, personal communications).

Incompatible controls did not demonstrate the increase in cell death associated with vegetative incompatibility - *het-c* incompatible heterokaryons have approximately 15% cell death (Jacobson *et al.*, 1998; Q. Xiang and S. Sarkar, personal communications); R1.21 + FGSC 5448 = 7% and R1.21 + MH27.131 = 5%. The reduced percentage of cell death in the mating-type incompatible controls may be explained with an escape of the incompatible phenotype. Mating-type associated vegetative incompatibility is usually noted with a severe growth restriction, which leads to a low mycelial mass. By waiting for mycelial material for cell death assessment (3 days), it is possible that the heterokaryon had gone through an escape to overcome the incompatible phenotype. Incompatible het tests with MH27 and MH27 derived strains forced with FGSC 5448 cell death ranged from 9-23%, despite an established Sup (mvi) phenotype.

Inconsistent cell death percentages may be due to a number of reasons. As mentioned above, only one trial per strain or test was completed, with an assessment of approximately 200 cells. More cell counts and additional test plates would need to be counted to determine significant results. The het tests constructed for the cell death assessment grew at a slower rate that previous het tests and the partial suppression of mating-type associated vegetative incompatibility was less evident. The reason for this observation is unclear, but additional tests and counts would verify if this is a trend with the use of the cellophane membrane. R1.21 and MH27.14 grew at a slightly slower rate than the other strains used in this trial of het tests, which may have contributed to the slowed het test growth rates. Overall, there was an increase in cell

death in MH27 strains forced with the opposite mating-type, but more data is required to determine if these results are statistically significant and to establish consistent controls.

3b.3.1.5 Partial diploid analyses with MH27, MH27.14 and MH27.45

Crossing translocation strains to normal sequence strains permits analysis of the co-existence of duplicated regions in the same nucleus and was one of the methods used to examine and identify many of the het loci (Perkins, 1975; Mylyk, 1975). Partial diploid analyses were completed by crossing MH27 and MH27 derived strains with translocation strains for the mating-type locus (LG I), R4.33 (*T (IL→IIR)* 39311; *trp-4 tol A*) and R4.35 (*T (IL→ IIR)* 39311; *ser-3 A*). If the MH27 mutation is a dominant suppressor of mating-type associated vegetative incompatibility and linked to or in *mat a*, all of the mating-type partial diploid progeny should result in a partially compatible appearance, despite presence of *mat A* and *mat a* in the same nucleus. Two translocation strains were used to assess the Sup (mvi) phenotype in *tol+* and *tol–* backgrounds.

One hundred twenty PDA and PDB progeny were picked to Vogel's minimal medium slants (Vogel, 1964) with 1x concentration of appropriate supplements. Sixty ascospores were picked per PDC, PDD, PDE and PDF crosses (PDA = MH27.14 x R4.35; PDB = MH27.14 x R2.33; PDC = MH27 x R4.35; PDD = MH27 x R4.33; PDE = MH27.45 x R4.35; PDF = MH27.45 x R4.33; PDI = I-1-51 x R4.35; PDJ = I-1-51 x R4.33). One hundred twenty-five progeny were picked from the control PDI and PDJ crosses. All viable progeny were screened for phenotype and growth on Vogel's minimal medium with the appropriate supplements. Strains were genotyped using BdeS medium plates (recipe in section **3a.2.2.6.3**) and mating-type

was determined with *fl a* (*mat*) and FGSC 7214 + a^{m1} (*mat A*) tester strains, as mentioned previously. Only the data for the crosses with the overall results for the relevant crosses with R4.35 are included in **Table 13** because the *tol-* phenotype was not evident in the I-1-51 control cross to R4.33 (data not shown). Because the *tol-* phenotype was not observed, the corresponding data was discarded due to doubt in validity of the *tol-* translocation strain.

Table 13. Partial diploid cross results

CROSS	0:A:a:A/a RATIOS	COMPATIBLE : INCOMPATIBLE RATIOS	GENOTYPE RATIOS
PDA = MH27.14 (*ad-8 lys-5 a*) x R4.35 (*T(I-II); ser-3 A*)	29:10:15:5 (~ 3:2:2:1)	22:8 (~ 3:1)	*lys-5* = 13 - : 13 + *ser-3* = 9 - : 17 + *ad-8* = 13 - : 13 +
PDC = MH27 (*pan-2 a*) x R4.35 (*T(I-II) ser-3 A*)	14:19:20:5 (~ 2:2:2:1)	34:12 (~ 3:1)	*pan-2* = 21 - : 21 + *ser-3* = 16 - : 26 +
PDE = MH27.45 (*pan-2 a*) x R4.35 (*T(I-II); ser-3 A*)	16:15:21:6 (~ 2:2:2:1)	30:14 (~ 2:1)	*pan-2* = 18 - : 20 + *ser-3* = 14 - : 24 +
PDI = I-1-51 (*ad-3A nic-2 a*) x R4.35 (*T(I-II); ser-3 A*)	7:9:12:4 (~ 2:2:3:1)	18:8 (~ 2:1)	*ad-3A* = 12 - : 12 + *nic-2* = 12 - : 12 + *ser-3* = 10 - : 14 +

Partial diploid analysis did not support the MH27 mutation as a dominant suppressor of mating-type associated vegetative incompatibility. In all classes, there was an approximately wild type ratio of ~2:1 of compatible (*mat A* or *mat a*) to incompatible partial diploid class progeny (*mat A/mat a*). Two progeny from two separate crosses were scored as *mat A/mat a* and displayed a wild type growth, which may represent a mating test error (data not shown). All other *mat A/mat a* progeny demonstrated an incompatible phenotype. Segregation of the markers was 1:1, with the exception of *ser-3* (column four of **Table 13**).

In general, the partial diploid analysis of the Sup (mvi) phenotype shows that the mutation cannot suppress mating-type associated vegetative incompatibility in a heterozygous

condition, indicating the partial suppression phenotype may be the result of a recessive mutation or a distortion in nuclear ratios in a heterokaryon. Nuclear ratio issues were not addressed in the partial diploid analysis.

3b.3.1.6 Complementation of the Sup (mvi) phenotype

Since the Sup (mvi) phenotype appeared to be linked to *mat a*, transformation of *mat a-1* into MH27 could determine if the mutation is in *mat a-1*, if there is complementation of the Sup (mvi) phenotype. If the mutation is outside *mat a-1*, there should be no restoration of mating-type vegetative incompatibility in the transformed MH27 strains. As well, native *mat a-1* vegetative incompatibility function in MH27 was assessed through a transformation of *mat A-1* into MH27.

3b.3.1.6.1 Transformation reduction assay

MH27 *mat a* vegetative incompatibility function was assessed using a transformation reduction assay. Transformations were completed with pCB1004 (control), *mat a-1* and *mat A-1* into MH27 and a^{m1} (*pan-2 a* and *ad-3B cyh-1 a^{m1}*, respectively; **Table 14**). The pCB1004 and a^{m1} transformations served as controls in determining if there was a complementation of the Sup (mvi) phenotype in the MH27 transformations.

Table 14. Transformation assay results using MH27 and a^{m1}

STRAIN	TRANSFORMATION TRANSFORMED FRAGMENT	NUMBER OF TRANSFORMANTS	CONCLUSION
a^{m1}	pCB1004	53	Compatible
MH27	pCB1004	>100	Compatible
a^{m1}	*mat a-1*	29	Compatible
MH27	*mat a-1*	>100	Compatible
a^{m1}	*mat A-1*	49	Compatible
MH27	*mat A-1*	2	Incompatible

A reduction in transformation efficiency was noted in MH27 with transformation of *mat A-1* (**Table 14**). Transformation reduction is an incompatible phenotype associated with the presence of alternative *het* alleles in the same nucleus, indicating that MH27 retains its *mat a-1* vegetative incompatibility function. a^{m1} transformations yielded a relatively low number of transformants, but this may be a reflection of the transformation efficiency of the spheroplasts. a^{m1} served as a control; a^{m1} by itself has no mating-type associated vegetative incompatibility and is sterile, indicating that *mat a-1* is non-functional. All other transformation assay results appeared to be compatible, with a significant number of transformants compared to the MH27 *mat A-1* transformation assay.

The conclusion from this assay is that the mutation causing the Sup (mvi) phenotype does not affect the vegetative incompatibility function of *mat a*.

3b.3.1.6.2 MH27 and a^{m1} transformant het tests

If the mutation causing the Sup (mvi) phenotype in MH27 is in *mat a-1*, transformation of a wild-type *mat a-1* into MH27 should restore mating-type vegetative incompatibility. Het tests with MH27 and a^{m1} transformation strains (described in detail in **3b.2.1.6.3**) and control

145

R1.21 and MH27.131 strains (*pyr-4 arg-5 pregc-2 trp-3 a* and *ad-8 lys-5 A* respectively) were completed with at least three different homokaryotic transformants, except for the single *mat A-1* transformant (**Table 15**). **Table 15** represents the average cumulative growth for the het tests over a three-day period. An average growth rate per day could not be determined because race tube analyses were completed in tubes with a maximum length of 13 cm and compatible heterokaryons usually grow this length in 2-3 days.

Table 15. Compilation of het test growth rates for MH27 and a^{m1} transformants

CONTROL STRAIN OR TRANSFORMANT (mating-type)	AVERAGE GROWTH ACCUMULATION OVER 3 DAYS (cm) ± STDEV* R1.21 (*mat a*)	AVERAGE GROWTH ACCUMULATION OVER 3 DAYS (cm) ±STDEV* MH27.131 (*mat A*)
I-1-51 (*mat a*)	11.05	
FGSC 5448 (*mat A*)		>13
MH27.131 (*mat A*)	1.25	
a^{m1} (pCB1004) (no mating--type)	7.4 ± 1.6	10.8 ± 2.2
MH27 (pCB1004) (*mat a*)	11.3 ± 2.1	4.5 ± 1.7
a^{m1} (*mat a-1*) (*mat a*)	11.1 ± 1.4	9.2 ± 3.3
MH27 (*mat a-1*) (*mat a*)	11.2 ± 1.6	5.75 ± 2.0
a^{m1} (*mat A-1*) (*mat A*)	10.6	11.8
MH27 (*mat A-1*) (*mat a + mat A*)	>13	6.0

*Note: Standard deviations are not included for het tests with less than 3 trials

Since there is high variability among the transformation strains (1.4 is the smallest standard deviation in **Table 15**), it is difficult to make conclusions about the complementation of the Sup (mvi) phenotype with *mat a-1*. The a^{m1} pCB1004 control transformant strains performed as expected, with a compatible phenotype with both R1.21 and MH27.131. The *mat a-1* construct transformed into a^{m1} did not appear to significantly complement the loss of vegetative incompatibility associated with the a^{m1} mutation, which may indicate that the construct is not fully functional. Similar results were seen with the transformation of *mat A-1* into a^{m1}.

In general, the average growth of the *mat a-1* transformed MH27 strains (5.75 ± 2.0 cm) is higher than the control transformation with pCB1004 (4.5 ± 1.7 cm), but the difference between these figures is not significant. The single MH27 (*mat A-1*) transformant tested appeared to function as *mat a*, with an average growth of 6.0 cm for the *mat A* het test versus >13 cm with the *mat a* het test. In this trial, only one transformant was tested and there is a high risk of escape, as *mat A/mat a* in the same nucleus leads to an incompatible phenotype.

The mating-types of the heterokaryons from the het tests were assessed to reveal the presence or absence of the mating-type function(s). All heterokaryons forced with MH27 transformant strains yielded *mat a* mating-type regardless of the forcing strain, while the a^{m1} controls gave a combination of *mat A*, *mat a* or both (**Table 16**).

Table 16. Het test mating-type assessment for MH27 and a^{m1} transformants

TRANSFORMANT	TRANS. #	MATING-TYPE OF HETEROKARYON	
		R1.21	MH27.131
a^{m1} (pCB1004)	1	*a*	*A*
	2	*A*	*A*
MH27 (pCB1004)	1	*a*	*a*
	2	*a*	*a*
a^{m1} (*mat a-1*)	1	*a*	*A*
	2	*a*	ND
MH27 (*mat a-1*)	1	*a*	*a*
	2	*a*	*a*
	3	*a*(few)	ND
	4	*a*	ND
a^{m1} (*mat A-1*)	1	*A*(few)/*a*	ND
	2	*A*(few)/*a*	*A*
MH27 (*mat A-1*)	1	*a*	*a*(few)

The results from the mating-type tests of the heterokaryons indicate that there may be a distortion in nuclear ratio, favoring the mutant nucleus. The reasons for this are unclear, but the same results were seen in het tests with untransformed MH27 derived strains.

Overall, there is no support that the mutation causing the Sup (mvi) phenotype is in *mat a-1* from the complementation assays, but new construct design and transformations (including other MH derived strains) would be required to differentiate whether the Sup (mvi) phenotype is in *mat a-1* or outside this gene.

3b.3.1.7 Sup (mvi) summary

A novel mutation in the strain MH27 causes partial suppression of mating-type associated vegetative incompatibility (Sup [mvi] phenotype). The mutation is linked to or in *mat a* and is marker independent. Het test growth rate analyses with MH27 derived strains show a normal growth rate, but may be dependant on a double mutation with *ncvip1*, as single Sup (mvi) mutants have a rapid decrease in growth rate after 4 or 5 days. This is demonstrated with a wild type growth rate in MH27.14 (*ad-8 lys-5 ncvip1* [?] *a*) versus the decrease in growth rate of MH27.11 (*ad-8 lys-5 a*) and MH27 (*pan-2 a*) when forced with a strain of the opposite mating-type. Cell death assays show an intermediate percentage of cell death when strains with the Sup (mvi) phenotype are forced with strains of the opposite mating-type, supporting the observation of partial suppression of mating-type vegetative incompatibility. Transformation of *mat A-1* into MH27 yielded a reduction in transformation efficiency, an indication of a functional *mat a* due to the incompatible response; *mat a-1* failed to restore mating-type vegetative incompatibility to MH27 when transformed strains were forced with strains of the opposite mating-type. Finally, the Sup (mvi) phenotype may cause suppression of incompatibility through a distortion of nuclear ratios by an unknown mechanism. More analysis is needed, including sequencing *mat a-1* in MH27, to determine the nature of the mutation.

3.4 DISCUSSION

TOL's involvement in the mating-type associated vegetative incompatibility response and the presence of two putative protein-protein interaction domains prompted a screen of a mycelial *N. crassa* cDNA library using the yeast two-hybrid system (Shiu, 2000). Two putative positive clones were identified as encoding putative proteins that interact with TOL (*pit* genes - Shiu, 2000 and this study). Sequence analysis of 7A10 and 7D8/*ncvip1* showed they encode 130 and 283 amino acids respectively. *ncvip1* was chosen for further analysis due to similarity with Vip1 in *S. pombe*, an uncharacterized protein. This project comprised the creation of mutants of *ncvip1* via RIP to determine the role of *ncvip1* in mating-type associated vegetative incompatibility. In the construction of *ncvip1* RIP mutants, an independent mutant showing partial suppression of mating-type associated vegetative incompatibility (MH27) was identified and characterized as part of this study.

3.4.1 Complications using the yeast two-hybrid system to identify clones encoding proteins that interact with TOL (*pit* genes)

The results of the cDNA library screen by Shiu (2000) demonstrate the potential problems with the library, the lack of true positive clones; two of the initial six "positive" clones encoded appropriate in-frame polypeptides (**Table 5**). This may indicate that interactions with TOL are transient or too weak to be detected by this method or that TOL does not interact with other proteins. Other considerations that must be made about TOL concern potential misfolding or mislocalization that may affect the ability of TOL to interact with other proteins, although TOL is predicted to be a nuclear protein.

The fact that only two positive clones were identified in the mycelial cDNA library screen indicates that there may be problems with the cDNA library used in the yeast two-hybrid screen. There are indications that the library has a large number of short cDNAs, low representation in the overall number of expressed sequences and a high level of false positives upon transformation into the yeast-two hybrid system (K. Borkovich; personal communications; P. Shiu and L. Glass, personal communications). For future studies, it may be necessary to use a different library for analysis or establish first if the protein of interest does in fact interact with other proteins in a tagged protein immunoprecipitation experiment. The caveat of both of these approaches is limitations of low levels of proteins expression; TOL is not expressed in high levels, as RT-PCR was required for detection (Shiu and Glass, 1999) which may add complications to an immunoprecipitation experiment.

3.4.2 *ncvip1*

3.4.2.1 Sequence analysis of *ncvip1*

With the identification of *ncvip1*, a cDNA clone that encoded a protein that interacted with TOL in the yeast two-hybrid system, further characterization was necessary. RFLP mapping localized *ncvip1* to the left arm of linkage group VI. Molecular characterization of the regulatory and processing consensus sequences are found in section **3a.3.2.1.3**. *ncvip1* encodes a putative 283 amino acid polypeptide, with a 63 bp intron (**Figure 33**; Bruchez *et al.*, 1993). Comparison with Vip1, a similar protein in *S. pombe* (257 aa; **Figure 27**), corresponds to the 283 amino acid polypeptide and there is a good consensus around the ATG of *ncvip1* compared to published consensus sequences (**Figure 33**; Edelmann and Staben, 1994). The true start can only

be determined using primer extension of labeled DNA primers to determine if the cDNA that was previously identified is in fact full length (Sambrook *et al.*, 1989).

Protein sequence analysis of NCVIP1 using two prediction algorithms (Pfam and SMART) identified an RNA recognition motif (RRM) in the N-terminal region of NCVIP1 (**Figure 34**). The RRM motif is found in a large variety of ribonucleoproteins (RNPs) that bind pre-mRNA, mRNA, pre-ribosomal RNA and small nuclear RNAs (reviewed in Kenan *et al.*, 1991; Burd and Dreyfuss, 1994). Because this domain is found in such a large variety of proteins with diverse roles and structure, very little can be hypothesized in reference to the significance of the domain in NCVIP1 beyond the likelihood that NCVIP1 will bind RNA. How this relates to a potential role in mating-type associated vegetative incompatibility remains to be seen, as even the role of *ncvip1* in this response is undetermined. An analysis of general patterns of *ncvip1* expression and phenotype in wild type and *tol-* backgrounds over a time course would be necessary before looking into the ability of NCVIP1 to bind RNA and the consequence of this function.

3.4.2.2 RIP analyses of *ncvip1*

RIP was the method of choice for mutational analysis of *ncvip1* since the cDNA was cloned and sequenced. As mentioned previously, *N. crassa* scans the genome for duplications during the sexual cycle, identifies them (presumably via pairing) and mutates both sequences through G:C to A:T transitions and alters the methylation pattern at the cytosine residues (Selker *et al.*, 1987; Selker 1990).

The initial *ncvip1* RIP cross had a bias against selection of *pan-2* progeny because pantothenic acid was not included in crossing media, as is a requirement of the *pan-2* mutation. Of the 47 hygromycin sensitive progeny screened, 17 progeny and the original transformed parent had methylation changes and hybridized to the pCB1004 vector, indicating maintenance of the vector sequence despite selection against the hygromycin resistant cassette (section **3a.3.2.3.2**). A study investigating the meiotic stability of the *hygR* sequence from *Escherichia coli* transformed into *Gibberella fujikuroi* (*Fusarium moliliforme*) found that single copy transformants had the expected 1:1 segregation in progeny of *hygR : hygS*. A skewed ratio of 1 *hygR : 2 hygS* was noted with multiple integrations of the *hygR* sequence. Some of the progeny also tested as *hygS*, but retained the introduced *hyg* sequence and passed it onto progeny that subsequently became *hygR* (Leslie and Dickman, 1991). This shows that methylation alterations may influence the expression of *hygR*; methylation is known to occur with the presence of duplicated sequence (Selker, 1990). How the reversion to *hygR* from *hygS* occurs is unclear. It also remains unclear if methylation of the gene can cause silencing in *N. crassa*.

MH27 (*pan-2 a*), one of the potentially *ncvip1* RIPed progeny with methylation changes, was sequenced and found to contain no changes in *ncvip1* resident DNA sequence. This may be an anomaly among the progeny with methylation changes, since this was the only *ncvip1* RIP cross progeny analyzed by sequence analysis. MH27 was subsequently taken through a cross with FGSC 7214 (*ad-8 lys-5 A*) to attempt to segregate the methylation changes associated with the ectopic copy from the resident copy and to increase the likelihood of RIP mutations. Of the 43 screened progeny in the second cross, 23 had methylation changes, following the Mendelian segregation pattern of 1:1. Again, these progeny were hybridized with the vector pCB1004 and all but one (MH27.69) was found to maintain the vector. This indicates that the plasmid

integration in the initial transformant was very close to the resident copy of *ncvip1*, linking the two copies. This event would explain why essentially all of the progeny hybridized to both the *ncvip1* and pCB1004 probes. Since one of 23 progeny with methylation changes segregated from vector sequence, this indicates a 4% linkage between the *ncvip1* locus and the ectopic copy, a separation of approximately 80 kb.

The key progeny generated in this study, with regards to *ncvip1*, is MH27.69 (*pan-2 A*). MH27.69 has methylation changes in *ncvip1* and no vector sequences when probed with *ncvip1* and pCB1004. MH27.69 was the only progeny generated that segregated the methylation changes from vector and is a putative *ncvip1* RIP progeny; the phenotype of MH27.69 appears wild type, indicating that *ncvip1* is non-essential.

The next step in characterizing the role of *ncvip1* is to sequence *ncvip1* in MH27.69 to see if it is a true null or partial-loss of function mutant. If MH27.69 is a *ncvip1* mutant, as supported with methylation changes in *ncvip1*, MH27.69 in future experiments should be crossed with a strain with markers on linkage group VI (where *ncvip1* is located). The progeny would be screened for methylation changes to select additional *ncvip1 mat A* and *mat a* mutants and to generate mutant strains with more markers for use in assessing *ncvip1* homokaryotic influence on mating-type vegetative incompatibility via het tests. This was the ultimate goal of the project and permits assessment of *ncvip1* mutations in a recessive condition. A cross with MH27.69 (*pan-2 ncvip1 A*) has been completed with 201 (*his-5 trp-4 tol; pan-2 a*) and now awaits assessment.

3.4.2.2.1 Explaining complications of RIP: methylation and silencing

One of the complications that arose in this study was the *de novo* methylation of the *ncvip1* sequence that obscured the changes in methylation RFLP patterns that may have been associated with RIP. It is possible that the introduction of the cDNA versus the genomic copy of *ncvip1* was recognized by the fungus as being foreign and caused the methylation of *ncvip1*. Methylation of transformed DNA is not uncommon in *N. crassa* and was noted in the transformation assays for RIP of *mat A-2* and *mat A-3* (L. Glass, personal communications; general reference - Selker *et al.*, 1993). In mammalian systems, methylation usually occurs at CpG sequences (Jaenisch and Jahner, 1984) and there is a direct correlation of methylation and gene silencing (X chromosome inactivation [Jaenisch *et al.*, 1998] and imprinting [Bartolomei and Tilghman, 1997]). This is in contrast to *N. crassa*, where cytosine methylation is less restricted (Selker *et al.*, 1993) and there is no direct evidence that methylation alone causes gene silencing.

From studies examining the effects of silencing mechanisms in *N. crassa* (quelling - Cogoni and Macino, 1999a and b, transvection - Aramayo and Metzenberg, 1996 and RIP – Selker *et al.*, 1987; Selker, 1990), pairing between duplicated gene sequences or RNA and DNA appears to have a more important role in silencing than methylation. As mentioned previously, RIP can cause gene silencing due to pairing of sequences resulting from duplications in the genomic sequence and silencing occurs with G:C to A:T transition mutations that appear in both copies of the DNA sequence (Selker *et al.*, 1987; Selker 1990). These mutations are also accompanied by alterations in the cytosine methylation patterns, but the role of methylation in gene regulation remains unclear. MIP (methylation induced premeiotically; Goyon and

Faugeron, 1989; Rossignol and Faugeron, 1994), a silencing process in *Ascobolus immersus*, has methylation changes associated with gene silencing, but not the DNA base changes.

Quelling (Cogoni and Macino, 1999a and b) is post transcriptional gene silencing (PGTS) mediated by transgenes, which requires RNA directed RNA polymerases and other RNA binding proteins. It is thought that quelling may serve as a means of silencing viral DNA or transposons (Cogoni and Macino, 1999a and b). This is interesting considering the RRM motif in NCVIP1, although there is no experimental evidence linking *ncvip1* to PTGS.

3.4.3 Sup (mvi) analyses and its role in mating-type associated vegetative incompatibility

3.4.3.1 Characterization of a novel mutation in strain MH27

Het tests with MH27 identified a partial suppression of mating-type associated vegetative incompatibility (**sup**pressor of **m**ating-type **v**egetative **i**ncompatibility = Sup [mvi] phenotype - **Figure 39**). Partial suppression was not evident using *het-c*, indicating that the Sup (mvi) phenotype was specific for mating-type and was unlikely to be due to cross feeding.

Initial characterization of the Sup (mvi) phenotype was accomplished through sequence analysis of the MH27 *ncvip1* locus, which showed that the *ncvip1* sequence was intact. Thus, the Sup (mvi) phenotype was caused by a novel mutation that worked in a semi-dominant fashion in a heterokaryon. MH27 was taken through a cross and progeny demonstrated a 1:1 segregation of the Sup (mvi) phenotype indicated that the mutation was heritable (**Table 9**). Genotyping of MH27 derived progeny showed that the phenotype segregated with *mat a*, with only one putative progeny (MH27.124) being *mat a* and not having the Sup (mvi) phenotype (**Table 9**). None of the *mat A* progeny had the Sup (mvi) phenotype (**Table 9**). This localized the gene causing the

Sup (mvi) phenotype to *mat a* or a gene tightly linked to *mat a*. *mat a-1* should be sequenced to determine if there are any mutations in *mat a* that could give rise to the Sup (mvi) phenotype. A *mat a-1* deletion strain (Δ216-220) has mating ability, but does not confer mating-type associated vegetative incompatibility (Philley and Staben, 1994). It may be that the mutation causing the Sup (mvi) phenotype occurs in this region, which could be easily identified by sequence analysis.

Genotyping of MH27 derived progeny also determined that the Sup (mvi) phenotype is marker independent because a variety of marker combinations were isolated (**Table 9**). It should be noted, however, that the *ad-8* and *lys-5* markers did not appear to segregate 1:1 as expected (18/59 *ad-8*; 17/59 *lys-5*). The reason for this bias is unclear, unless it is skewed toward the mutant nuclei, which would support the theory that the Sup (mvi) phenotype is due to a distortion in nuclear ratios.

Partial diploid analysis did not demonstrate a partial suppression of mating-type associated vegetative incompatibility. MH27 and MH27 derived strains with the Sup (mvi) phenotype were crossed to strains containing a translocation of the mating-type locus (LG I) to LG II; there was no partial suppression of mating-type vegetative incompatibility in the *mat A / mat a* partial diploid class of progeny (**Table 13**). If the gene causing the Sup (mvi) phenotype was a dominant or semi-dominant suppressor and linked to or in *mat a*, all partial diploid progeny should appear to grow as wild-type or similar to the *mat A* + MH27 (*mat a*) heterokaryons showing partial suppression.

To determine if the Sup (mvi) phenotype could be complemented by *mat a-1* and thus support that the Sup (mvi) phenotype was due to a mutation in *mat a-1*, transformations of a *mat a-1* construct were completed in MH27. Results were inconclusive (**Table 15**), leaving the nature of the mutation causing the Sup (mvi) phenotype unsettled. Another transformation, this

time with *mat A-1* into MH27, showed a reduction in transformation efficiency (an incompatible reaction), indicating a functional MAT a-1 with respect to vegetative incompatibility.

The Sup (mvi) mutation may cause a nuclear distortion ratio or alter the protein stability of MAT a-1, both of which could cause a partial suppression phenotype in heterokaryons but not in partial diploids or transformants. A distortion in nuclear ratio is supported with the MH27 derived progeny and the bias against *ad-8 lys-5* markers (the markers of FGSC 7214; **Table 9**). All heterokaryons forced with transformed MH27 (*mat a-1* and *mat A-1*) mated as *mat a*, another bias toward the markers of the mutant nuclei (data not shown). The overall finding is that the Sup (mvi) phenotype is not expressed when *mat A-1* and *mat a-1* are transcribed and translated in the same nucleus, but is expressed when the gene products are expressed in different nuclei, as in a heterokaryon. Exploration of the nuclear ratio issue may be possible by creating tagged *mat A-1* and *mat a-1* constructs and visualizing location with immunofluorescent methods. This would permit analysis of the location of *mat A-1* and *mat a-1* and the approximate nuclear ratios in a heterokaryon. It would also be evident if in fact *mat A* gene products are being suppressed in the MH27 background.

The same phenotype of suppression of mating-type incompatibility in heterokaryons but not transformants was observed by Saupe *et al.* (1996a) in generating novel mutants in *mat A-1*. A^{m64} has a frame-shift mutation at amino acid position 111 and confers vegetative incompatibility in transformation assays, but suppresses vegetative incompatibility in heterokaryons (Saupe *et al.*, 1996a). As mentioned in the introduction, MAT A-1 and MAT a-1 are likely transcriptional regulators (Glass *et al.*, 1990; Staben and Yanofsky, 1990). One proposed hypothesis is that the gene products are differentially targeted to the nucleus of origin, limiting the interaction in a common cytoplasm or having a more directed function in the nucleus of origin. Mutations that

disrupt this targeting would not trigger the incompatible response in a heterokaryon because the gene products would not have the opportunity to interact in the nucleus. If the genes are located in the same nucleus, as in the case with transformants and partial diploids, there could be an interaction leading to vegetative incompatibility. There are indications that the sequence 163-293 bp in *mat A-1* may modify or enhance mating-type associated vegetative incompatibility. This is supported by the observation that there is a larger reduction in transformation efficiencies with a complete *mat A-1* sequence as compared to the truncated A^{m54} and A^{m64} alleles (Saupe *et al.*, 1996a). An interesting experiment would be to transform A^{m64} into MH27 *mat a* to see if in a double mutant background for heterokaryon suppression of mating-type vegetative incompatibility, transformants can still confer normal mating-type associated vegetative incompatibility function.

3.4.3.2 Het tests with MH27 and MH27 derived strains

Race tube growth rate analysis of het tests with MH27 strains may indicate a synergistic effect of mutations in *ncvip1* and the mutation causing the Sup (mvi) phenotype. There may also be an influence of temperature. MH27 (*pan-2 a*) het tests with FGSC 5448 (*ad-3A; inl A*) showed a high variability between trials and exhibited a dramatic decrease in growth rate after approximately four days (**Table 11** and data not shown). The same phenomenon was also observed in het tests with MH27.11 (*ad-8 lys-8 a*) and FGSC 5448, indicating that the slowed growth rate is not marker dependent. What is known is that both of these strains contain an intact version of *ncvip1* (MH27 verified by sequence analysis and the lack of methylation changes when probed with *ncvip1* in MH27.11). In contrast, MH27.14 (*ad-8 lys-5 a*) does not

experience the slowed growth rate when forced with FGSC 5448 (**Table 11**). In a hybridization to *ncvip1*, MH27.14 displayed methylation changes, which may be associated with RIP of *ncvip1*. From these data, it appears that there may be an interaction between *ncvip1* and the gene controlling the Sup (mvi) phenotype. Sequence analysis of *ncvip1* in MH27.14 will need to be completed to determine if in fact MH27.14 is a double mutant with the mutation causing the Sup (mvi) phenotype. Additional het tests with other progeny and progeny with different nutritional markers may provide more conclusive evidence of synergism. An appropriate control and comparison would be the use of MH27.69, the one strain that has methylation changes in *ncvip1* that segregated from the vector sequence. This strain is putatively mutated in *ncvip1* and does not have the Sup (mvi) phenotype.

The influence of temperature on the Sup (mvi) phenotype is unclear. Contradictory results with MH27 found that the slowing in growth rate when a heterokaryon was forced with a strain of the opposite mating-type at room temperature was not apparent at 10°C (**Table 11**). Similar het tests with MH27.14 demonstrated a compatible phenotype at room temperature, but an incompatible phenotype at 10°C (**Table 11**). Putative interaction between *ncvip1* and the gene causing the Sup (mvi) phenotype at 10°C can not be postulated because there is no data with MH27.11 for comparison. The variation in phenotypes between MH27 and MH27.14 at room temperature versus 10°C may point towards issues of protein stability. There may be differences in regulation of one or more of the proteins involved in recognition or regulation of the mating-type vegetative incompatibility response. Mutations may cause changes that affect protein stability, regulation of protein products or signaling or transport of proteins in and out of the nucleus. More experimental results are needed to establish the possible synergism before any solid conclusions can be made.

3.4.4 Ascospore viability

Viability of ascospores limited the number of *ncvip1* RIP progeny available for screening of methylation changes and consequently RIP (data not shown due to variable germination conditions). Over time, the viability of ascospores declined, although there may be bias in the number of germinated ascospores due to periods of heat shock at 60°C extending past 30 minutes, as required to stimulate ascospore germination. It was also noticed that there is a greater viability and germination of ascospores germinated in a small aliquot of sterile, distilled water and then spread on a plate versus heat shocking the entire plate. As well, there was a better germination rate if ascospores were heat shocked for 20-25 minutes versus longer periods of time. The reason for the low viability of ascospores is unclear, as wild type control crosses also exhibited this phenotype.

3.4.5 Summary of future directions

Included here is a list of the immediate issues that should be addressed:

ncvip1

1) Sequence *ncvip1* in MH27.14 and MH27.69 to determine if they are RIPed.

2) Cross MH27.69 to introduce additional markers for het tests and to isolate *mat A* and *mat a* progeny; assess homokaryotic condition of *ncvip1* mutations on mating-type associated vegetative incompatibility.

3) Once *ncvip1* mutants have been identified, create double mutants with *tol* to assess potential synergistic effects in mating-type vegetative incompatibility.

<u>Sup (mvi) phenotype</u>

1) Sequence *mat a-1* in MH27 to determine if *mat a-1* DNA sequence is intact.

2) If the mutation is in another location, potential Sup (mvi) genes could be cloned through PCR using primers derived from putative ORF regions determined in the annotated genome. Once cloned, directed mutagenesis and deletion analyses can be completed to recreate partial suppression of mating-type associated vegetative incompatibility in heterokaryons.

3) Examine the role of mutation causing the Sup (mvi) phenotype in relation to *tol* and *ncvip1* through double mutant analyses.

4. CONCLUSION

Current understanding of the role of mating-type genes in vegetative incompatibility in *Neurospora crassa* is a result of cloning and characterization of these genes and their gene products (Glass *et al.*, 1990; Staben and Yanofsky, 1990; Ferreira *et al.*, 1996 and 1998). MAT A-1 and MAT a-1 interact in the yeast two-hybrid system and this interaction is thought to mediate the vegetative incompatibility response (Badgett and Staben, 1999). Although a mediator for the mating-type vegetative incompatibility response has been identified (*tol*; Newmeyer, 1970) and recently cloned (Shiu and Glass, 1999), little is known about the recognition and regulation of the pathway causing mating-type associated vegetative incompatibility. Homozygous *tol* mutants suppress mating-type vegetative incompatibility in heterokaryons.

There is a large prevalence of TOL-like sequences in the *N. crassa* genome and similarity is limited to a region termed the SET domain. This region is also present in HET-6 (*N. crassa*) and HET-E (*P. anserina*). Computational analyses have shown similarity of one of the regions in the SET domain (aa 377-343) to Mms21p, a protein involved in DNA repair in *S. cerevisiae*. The significance of this similarity and the existence of multiple TOL-like sequences in the *N. crassa* genome remains unclear.

Proteins that interact with TOL (encoded by *pit* genes) may help deduce the mechanism of *tol* function in *N. crassa* and how it is involved in regulating the vegetative incompatibility response with respect to mating type. Characterization of *ncvip1*, a putative *pit* gene, was the focus on this study. Unfortunately, mutational characterization of *ncvip1* is incomplete and one strain, MH27.69 may hold the key. This strain may be RIPed in *ncvip1*, but verification through sequence analysis is required.

162

A novel mutation causing the **sup**pression of **m**ating-type **v**egetative **i**ncompatibility (Sup [mvi]) phenotype arose in MH27 and the gene associated with this phenotype appears to be linked to or in *mat a*. Characterization of strains with the Sup (mvi) phenotype was difficult because of the potential for mutations caused by RIP in *ncvip1*. Some of the Sup (mvi) strains in hybridizations with *ncvip1* did not demonstrate methylation changes in *ncvip1* (which are associated with RIP), providing the tools for examination and characterization of the gene causing the Sup (mvi) phenotype. Complementation assays and partial diploid analysis did not shed light on the nature of the mutation, but it appears that some of the MAT a-1 vegetative incompatibility function remains intact because there was a reduction in transformation efficiency when MH27 was transformed with a *mat A-1* construct. This was also observed in *mat A* mutants, A^{m54} and A^{m64}, that showed a partial suppression of mating-type vegetative incompatibility in heterokaryons but not in transformation assays (Saupe *et al.*, 1996a). This may indicate that there is a preferential targeting of mating-type gene products to the nucleus of origin, but there is no experimental evidence to support this claim.

In conclusion, the role of *ncvip1* in mating-type associated vegetative incompatibility is unclear and the strain MH27.69 may provide the tools to assess if *ncvip1* is involved in this response. The novel mutation causing the Sup (mvi) phenotype may be an additional factor, but it must be determined whether or not it represents a new locus or is a mutation in *mat a-1*, a known component in mating-type associated vegetative incompatibility. Strains such as MH27.11 that have the Sup (mvi) phenotype but do not have methylation changes in *ncvip1* may provide insight into the role and function of this potentially novel component.

6. REFERENCES

Adam S.A., Nakagawa T., Swanson M.S., Woodruff T.K. and Dreyfuss G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol. Cell. Biol. **6(8):** 2932-2943.

Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology **215:** 403-410.

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

Anwar M.M., Croft J.H. and Dales R.B.G. (1993) Analysis of heterokaryon incompatibility between heterokaryon-compatibility (h-c) groups R and GL provides evidence that at least eight *het* loci control somatic incompatibility in *Aspergillus nidulans*. J. Gen. Microbiol. **139:** 1599-1603.

Aramayo R. and Metzenberg R.L. (1996) Meiotic transvection in fungi. Cell **86(1):** 103-113.

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

Badgett T.C. and Staben C. (1999) Interaction between and transactivation by mating type polypeptides of *Neurospora crassa*. Fungal Genet. Newsl. **46S:** 73.

Bartolomei M.S. and Tilghman S.M. (1997) Genomic imprinting in mammals. Annu. Rev. Genet. **31:** 493-525.

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

Becker W.M. and Deamer D.W. (1991) The World of the Cell, second edition. The Benjamin/Cummings Publishing Company, Inc. Menlo Park, CA. pp.757-760.

Bégueret J., Turcq B. and Clavé C. (1994) Vegetative incompatibility in filamentous fungi: *het* genes begin to talk. T.I.G. **10(12):** 441-447.

Bhuiyan M.K.A. and Arai K. (1993) Physiological factors affecting hyphal growth and function of *Rhizoctonia oryzae*. Trans. Mycol. Soc. Jpn. **32:** 389-397.

Birney E., Kumar S. and Krainer A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucl. Acids Res. **21(25):** 5803-5816.

Bistis G.N. (1981) Chemotropic interactions between trichogynes and conidia of opposite mating-type in *Neurospora crassa*. Mycologia **73**: 959-975.

Bistis G.N. (1994) Retardation of the growth of transplanted apothecia: a manifestation of vegetative incompatibility in *Ascobolus sterocorarius* (Bull.) Schröt. Exp. Mycol. **18**: 103-110.

Bobrowicz P. and Ebbole D.J. (2001) Components and function of the *Fus3/Kss1*-related MAP kinase pathway in *Neurospora crassa*. XXI Fungal Genetics Conference Asilomar, CA (abstract).

Boucherie H. and Bernet J. (1980) Protoplasmic incompatibility in *Podospora anserina*: a possible function for incompatibility genes. Genetics **96**: 399-411.

Boucherie H., Dupont C.H. and Bernet J. (1981) Polypeptide synthesis during protoplasmic incompatibility in the fungus *Podospora anserina*. Biochem. Biophys. Acta **653**: 18-26.

Brockman H.E. and deSerres F.J. (1962) Viability of Neurospora conidia from stock cultures on silica gel. Neurospora Newsl. **1**: 8-9.

Bruchez J.J.P., Eberle J. and Russo V.E.A. (1993) Minireview: Regulatory sequences in the transcription of *Neurospora crassa* genes: CAAT box, TATA box, introns, poly (A) tail formation sequences. Fungal Genet. Newsl. **40**: 89-96.

Buller A.H.R. (1933) *Researches on Fungi*. London: Longman. **5**: 416pp.

Burd C.G. and Dreyfuss G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science **265**: 615-621.

Carroll A.M., Sweigard J.A. and Valent B. (1994) Improved vectors for selecting resistance to hygromycin. Fungal Genet. Newsl. **41**: 22.

Caten C.E. (1972) Vegetative incompatibility and cytoplasmic infection in fungi. J. Gen. Microbiol. **72**: 221-229.

Cogoni C. and Macino G. (1999a) Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. Science **286**: 2342-2344.

Cogoni C. and Macino G. (1999b) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. Nature **399**: 166-169.

Coustou V., Deleu C., Saupe S. and Bégueret J. (1997) The protein product of the *het-s* heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. Proc. Natl. Acad. Sci. U.S.A. **94**: 9773-9778.

Dales R.B.G. and Croft J.H. (1983) A chromosome assay method for the detection of heterokaryon incompatibility (*het*) genes operating between members of different heterokaryon compatibility (h-c) groups in *Aspergillus nidulans*. J. Gen. Microbiol. **129:** 3643-3649.

Davis R.H. (2000) Neurospora: Contributions of a model organism. Oxford University Press Inc. New York.

Debets A.J.M. and Griffiths A.J.F. (1998) Polymorphisms of *het*-genes prevents resource plundering in *Neurospora crassa*. Mycol. Res. **102:** 1343-1349.

Debets F., Yang X. and Griffiths A.J.F. (1994) Vegetative incompatibility in *Neurospora*: its effect on horizontal transfer of mitochondrial plasmids and senescence in natural populations. Curr. Genet. **26:** 113-119.

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

van Diepeningen A.D., Debets A.J.M. and Hoekstra R.F. (1997) Heterokaryon incompatibility blocks virus transfer among natural isolates of black *Aspergilli*. Curr. Genet. **32:** 209-217.

van Diepeningen A.D., Debets A.J. and Hockstra R.F. (1998) Intra- and interspecies virus transfer in Aspergilli via protoplast fusion. Fungal Genet. Biol. **25(3):** 171-180. abstract

Ebbole D. and Sachs M.S. (1990) A rapid and simple method for isolation of Neurospora crassa homokaryons using microconidia. Fungal Genet. Newsl. **37:** 17-18.

Edelmann S.E. and Staben C. (1994) A statistical analysis of sequence features within genes from *Neurospora crassa*. Exper. Mycol. **18:** 70-81.

Espagne E., Balhadère P., Bégueret J. and Turcq B. (1997) Reactivity in vegetative incompatibility of the HET-E protein of the fungus Podospora anserina is dependent on GTP-binding activity and a WD40 repeated domain. Mol. Gen. Genet. **256:** 620-627.

Featherstone C. and Jackson S.P. (1999) DNA double-strand break repair. Curr. Biol. **9:** R759-761.

Ferreira A.V.-B., An Z., Metzenberg R.L. and Glass N.L. (1998) Characterization of *mat A-2*, *mat A-3* and Δ*matA* mating-type mutants of *Neurospora crassa*. Genetics **148:** 1069-1079.

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

Fincham J.R.S. (1989) Transformation in fungi. Microbiol. Revs. **53:** 148-70.

Fincham J.R.S., Connerton I.F., Notarianni E. and Harrington K. (1989) Premeiotic disruption of duplicated and triplicated copies of the *Neurospora crassa am* glutamate dehydrogenase gene. Curr. Genet. **15:** 327-334.

Fischer-Parton S., Parton R.M., Hickey P.C., Dijksterhuis J., Atkinson H.A. and Read N.D. (2000) Confocal microscopy of FM4-64 as a tool for analysing endocytosis and vesicle trafficking in living fungal hyphae. J. Microscopy **198(3):** 246-259.

Gaff D.L. and Okong'O-Ogola O. (1971) The use of non-permeating pigments for testing the survival of cells. J. Exp. Bot. **22:** 756-758.

Garnjobst L. (1953) Genetic control of heterocaryosis in Neurospora crassa. Am. J. Bot. **40:** 607-614.

Garnjobst L. and Wilson J.F. (1956) Heterocaryosis and protoplasmic incompatibility in *Neurospora crassa*. Proc. Natl. Acad. Sci. U.S.A. **42:** 613-618.

Glass N.L., Jacobson D.J. and Shiu P.K.T. (2000) The genetics of hyphal fusion and post fusion recognition in filamentous ascomycete fungi. Annu. Rev. Genet. **34:** 165-86.

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

Glass N.L. and Kuldau G.A. (1992) Mating type and vegetative incompatibility in filamentous ascomycetes. Annu. Rev. Phytopathol. **30:** 201-24.

Glass N.L. and Lee L. (1992) Isolation of *Neurospora crassa A* mating type mutants by Repeat Induced Point (RIP) mutation. Genetics **132:** 125-133.

Glass N.L. and Nelson M.A. (1994) Mating-type genes in mycelial ascomycetes - in *The Mycota I: Growth, Differentiation and Sexuality*. Eds: Wessels/Meinhardt. Springer-Verlag; Berlin, Heidelberg.

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

Gooday G.W. (1975) Chemotaxis and chemotropism in fungi and algae. In *Primitive sensory and communication systems: The taxes and tropisms of microorganisms and cells*. Ed. M.J. Carlile. London: Academic. pp. 155-204.

Goyon C. and Faugeron G. (1989) Targeted transformation of Ascobolous immersus and de novo methylation of the resulting DNA sequences. Mol. Cell. Biol. **9:** 2818-2827.

Graïa F., Lespinet O., Rimbault B., Dequard-Chablat M., Coppin E. and Picard M. (2001) Genome quality control: RIP (repeat-induced point mutation) comes to Podospora. Mol. Microbiol. **40(3):** 586-595.

Griffiths A.J.F. (1982) Null mutants of the *A* and *a* mating type alleles of *Neurospora crassa*. Can. J. Genet. Cytol. **24:** 167-176.

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

Grosschedl R., Giese K. and Pagel J. (1994) HMG protein domains: architectural elements in the assembly of nucleoprotein structures. T.I.G. **10:** 94-99.

Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557-580.

Harkness T.A.A., Metzenberg R.L., Schneider H., Lill R., Neupert W. and Nargang F.E. (1994) Inactivation of the Neurospora crassa gene encoding the mitochondrial protein import receptor MOM19 by the technique of "sheltered RIP". Genetics **136:** 107-118.

Hickey P.C., Jacobson D.J., Read N.D. and Glass N.L. (2001) Live cell imaging of hyphal fusions in growing *Neurospora* colonies. XXI Fungal Genetics Conference Asilomar, CA (abstract).

Hoffman D.W., Query C.C., Golden B.L., White S.W. and Keene K.D. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. Proc. Natl. Acad. Sci. **88(6):** 2495-2499.

Howard-Flanders P. and Theriot L. (1966). Mutants of *Escherichia coli* K-12 defective in DNA repair and in genetic recombination. Genetics. **53(6):**7-50.

Jacobson D.J. (1992) Control of mating type heterokaryon incompatibility by the *tol* gene in *Neurospora crassa* and *N. tetrasperma*. Genome **35:** 347-353.

Jacobson D.J., Barton M.M., Dettman J.R., Hiltz M.D., Powell A.J., Saenz G.S., Hirsh J.C., Taylor J. W., Glass N.L. and Natvig D.O. (2001) *Neurospora* in western North America: a model system in the backyard. XXI Fungal Genetics Conference Asilomar, CA (abstract).

Jacobson D.J., Beurkens K. and Klomparens K.L. (1998) Microscopic and ultrastructural examination of vegetative incompatibility in partial diploids heterozygous at *het* loci in *Neurospora crassa*. Fungal Genet. Biol. **23:** 45-56.

Jaenisch R., Beard C., Lee J.T., Marahrens Y. and Panning B. (1998). Mammalian X chromosome inactivation. Novartis Foundation Symposium **214:** 200-209.

Jaenisch R. and Jahner D. (1984) Methylation, expression and chromosomal position of genes in mammals. Biochim. Biophys. Acta. **782:** 1-9.

Johnson T.E. (1978) Isolation and characterization of perithecial development mutants in Neurospora. Genetics **88:** 27-47.

Karran P. (2000) DNA double strand break repair in mammalian cells. Curr. Opin. Genet. Dev. **10:** 144-150.

Keene J.D. (1996) RNA recognition by autoantigens and autoantibodies. Mol. Biol. Rev. **23:** 173-181.

Kenan D.J., Query C.C. and Keene J.D. (1991) RNA recognition: towards identifying determinants of specificity. T.I.B.S. **16:** 214-220.

Kothe G.O. and Free S.J. (1998) The isolation and characterization of *nrc-1* and *nrc-2*, two genes encoding protein kinases that control growth and development in *Neurospora crassa*. Genetics **149:** 117-130.

Kwon K.J. and Raper K.B. (1967) Heterokaryon formation and genetic analyses of color mutants on *Aspergillus heterothallicus*. Amer. J. Bot. **54 (1):** 49-60.

Labarère J., Bégeuret and Bernet J. (1974) Incompatibility in *Podospora anserina*: comparative properties of the antagonistic cytoplasmic factors of a nonallelic system. J. Bacteriol. **120:** 854-860.

Leslie J.F. and Dickman M.B. (1991) Fate of DNA encoding hygromycin resistance after meiosis in transformed strains of *Gibberella fujikuroi* (*Fusarium moniliforme*). Appl. Environ. Microbiol. **57(5):** 1423-1429.

Leslie J.F and Yamashiro C.T. (1997) Effects of the *tol* mutation on allelic interactions at *het* loci in *Neurospora crassa*. Genome **40:** 834-840.

Lindahl T., Kerran P. and Wood R.D. (1997) DNA excision repair pathways. Curr. Opin. Genet. Dev. **7:** 158-169.

Marek S.M., Wu J., Glass N.L., Gilchrist D.G. and Bostock R.M. (1998) Programmed cell death in fungi: Heterokaryon incompatibility involves nuclear DNA degradation. Phytopathology **88:** A58.

Margolin B.S., Freitag M. and Selker E.U. (1997) Improved plamids for gene targeting at the *his-3* locus of *Neurospora crassa* by electroporation. Fungal Genet. Newsl. **44:** 34-36.

May G., Shaw F., Badrane H. and Vekemans X. (1999) The signature of balancing selection: fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. U.S.A. **96:** 9172-9177.

Metzenberg R.L. and Glass N.L. (1990) Mating type and mating strategies in *Neurospora*. BioEssays **12(2):** 53-59.

Metzenberg R.L. and Grotelueschen J.S. (1992) Disruption of essential genes in Neurospora by RIP. Fungal Genet. Newsl. **39:** 37-49.

Metzenberg R.L., Stevens J.N., Selker E.U. and Morzycka-Wroblewska E. (1984) A method for finding the genetic map position of cloned DNA fragments. Neuro. Newsl. **31:** 35-39.

Metzenberg R.L., Baisch T.J. and Stevens J.N. (1988) An easy method for preparing Neurospora DNA. Update from Neurospora Newsl. 28:20 and 29:27 (1981 and 1982).

Milgroom M.G. and Cortesi P. (1999) Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. Proc. Natl. Acad. Sci. U.S.A. **96:** 10518-10523.

Modrich P. (1991) Mechanisms and biological effects of mismatch repair. Ann. Rev. Genet. **25:** 229-253.

Montelone B.A. and Koelliker K.J. (1995) Interactions among mutations affecting spontaneous mutation, mitotic recombination, and DNA repair in yeast. Curr. Genet. **27:** 102-109.

Montelone B.A., Prakash S. and Prakash L. (1981) Hyper-recombination and mutator effects of the *mms9-1*, *mms13-1*, and *mms21-1* mutations in *Saccharomyces cerevisiae*. Curr. Genet. **4:** 223-232.

Mylyk O.M. (1975) Heterokaryon incompatibility genes in Neurospora crassa detected using duplication-producing chromosome rearrangements. Genetics **80:** 107-124.

Mylyk O.M. (1976) Heteromorphism for heterokaryon incompatibility genes in natural populations of *Neurospora crassa*. Genetics **83:** 275-284.

Nagai K., Oubridge C., Jessen T.H., Li J. and Evans P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature **348:** 515-520.

Nelson M.A. (1996) Mating systems in ascomycetes: a romp in the sac. T.I.G. **12(2):** 69-74.

Nelson M.A., Kang S., Braun E.L., Crawford M.E., Dolan P.L., Leonard P.M., Mitchell J., Armijo A.M., Bean L., Blueyes E., Cushing T., Errett A., Fleharty M., Gorman M., Judson K., Miller R., Ortega J., Pavlova I., Perea J., Todisco S., Trujillo R., Valentine J., Wells, A., Werner-Washburne M., Yazzie S. and Natvig D.O. (1997) Expressed sequences from conidial, mycelial, and sexual stages of *Neurospora crassa*. Fungal Genet. Biol. **21:** 348-363.

Nelson M.A., Crawford M.E. and Natvig D.O. (1998) Restriction polymorphism maps of Neurospora crassa: 1998 update. Fungal Genet. Newsl. **45:** 44-54.

Newmeyer D. (1970) A suppressor of the heterokaryon-incompatibility associated with mating type in *Neurospora crassa*. Can. J. Genet. Cytol. **12:** 914-926.

Oettinger M.A. (1999) V(J) recombination: on the cutting edge. Curr. Opin. Cell Biol. **11:** 325-329.

Pandit A. and Maheshwari R. (1996) Brief Note: A demonstration of the role of *het* gene in heterokaryon formation in Neurospora under simulated field conditions. Fung. Genet. Newsl. **20:** 99-102.

Perkins D.D. (1975) The use of duplication-generating rearrangements for studying heterokaryon incompatibility genes in Neurospora. Genetics **80:** 87-105.

Perkins D.D., Radford A. and Sachs M.S. (2001) *The Neurospora compendium: Chromosomal loci*. Academic Press, San Francisco U.S.A.

Philley M. and Staben C. (1994) Functional analyses of the *Neurospora crassa* MT a-1 mating type polypeptide. Genetics **137:** 715-722.

Pöggeler S. and Kück U. (2000) Comparative analysis of the mating-type loci from *Neurospora crassa* and *Sordaria macrospora*: identification of novel transcriptionally expressed ORFs. Mol. Gen. Genet. **263:** 293-301

Pöggeler S., Risch S., Kück U. and Osiewacz H.D. (1997) Mating-type genes from the homothallic fungus *Sordaria macrospora* are functionally expressed in a heterothallic ascomycete. Genetics **147:** 567-580.

Prakash S. and Prakash L. (1977) Increased spontaneous mitotic segregation in MMS-sensitive mutants of *Saccharomyces cerevisiae*. Genetics **87:** 229-236.

Raju N.B. (1980) Meiosis and ascospore genesis in *Neurospora*. Eur. J. Cell. Biol. **23:** 208-223.

Raju N.B. (1992) Genetic control of the sexual cycle in *Neurospora*. Mycol. Res. **96(4):** 241-262.

Rossignol J.-L. and Faugeron G. (1994) Gene inactivation triggered by recognition between DNA repeats. Experientia **50:** 307-317

Ryan F.J., Beadle G.W. and Tatem F.L. (1943) The tube method of measuring the growth rate of Neurospora. Amer. J. Bot. **30:** 784-799.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) *Molecular cloning – a laboratory manual* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring NY.

Saraste A. and Pulkki K. (2000) Morphological and biochemical hallmarks of apoptosis. Cardiovasc. Res. **45(3):** 528-537.

Saupe S.J. (2000) Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. Micro. Molec. Biol. Rev. **64(3):** 489-502.

Saupe S.J., Clavé C., Sabourin M. and Bégueret J. (2000) Characterization of *hch*, the *Podospora anserina* homolog of the *het-c* heterokaryon incompatibility gene of *Neurospora crassa*. Curr. Genet. **38:** 39-47.

Saupe S., Descamps C., Turcq B. and Bégueret J. (1994) Inactivation of the *Podospora anserina* vegetative incompatibility locus *het-c*, whose product resembles a glycolipid transfer protein, drastically impairs ascospore production. Proc. Natl. Acad. Sci. U.S.A. **91:** 5927-5931.

Saupe S.J. and Glass N.L. (1997) Allelic specificity at the *het-c* heterokaryon incompatibility locus of *Neurospora crassa* is determined by a highly variable domain. Genetics **146:** 1299-1309.

Saupe S.J., Kuldau G.A., Smith M.L. and Glass N.L. (1996b) The product of the *het-C* heterokaryon incompatibility gene of *Neurospora crassa* has characteristics of a glycine-rich cell wall protein. Genetics **143:** 1589-1600.

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

Saupe S., Turcq B. and Bégueret J. (1995) Sequence diversity and unusual variability at the *het-c* locus involved in vegetative incompatibility in the fungus *Podospora anserina*. Curr. Genet. **27:** 466-471.

Schweizer M., Case M.E., Dykstra C.C., Giles N.H. and Kushner S.R. (1981) Identification and characterization of recombinant plasmids carrying the complete *qa* gene cluster from *Neurospora crassa* including the *qa-1*$^+$ regulatory gene. Genetics **78(8):** 5086-5090.

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

Selker E.U., Cambareri E.B., Jensen B.C. and Haack K.R. (1987) Rearrangement of duplicated DNA in specialized cells of Neurospora. Cell **51:** 741-752.

Selker E.U., Richardson G.A., Garrett-Engele P.W., Singer M.J. and Miao V. (1993) Dissection of the signal for DNA methylation in the zeta-eta region of *Neurospora*. Cold Spring Harb. Symp. Quant. Biol. **58:**323-329.

Shear C.L. and 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.

Sheikh M.S. and Fornace A.J. (2000) Role of p53 family members in apoptosis. J. Cell. Phys. **182:** 171-181.

Shiu P.K.-T. (2000) Mating-type associated vegetative incompatibility in *Neurospora crassa*. Thesis. University of British Columbia.

Shiu P.K.T. and Glass N.L. (1999) Molecular characterization of *tol*, a mediator of mating-type-associated vegetative incompatibility in *Neurospora crassa*. Genetics **151**: 545-555.

Shiu P.K.T. and Glass N.L. (2000) Cell and nuclear recognition mechanisms mediated by mating-type in filamentous ascomycetes. Curr. Opin. Micro. **3**: 183-188.

Smith M.L., Hubbard S.P., Jacobson D.J., Micali O.C. and Glass N.L. (2000a) An osmotic-remedial temperature-sensitive mutation in the allosteric activity site of nucleotide reductase in *Neurospora crassa*. Mol. Gen. Genet. **262**:1022-1035.

Smith M.L., Micali O.C., Hubbard S.P., Mir-Rashed N., Jacobson D.J. and Glass N.L. (2000b) Vegetative incompatibility in the *het-6* region of *Neurospora crassa* is mediated by two linked genes. Genetics **155**: 1095-1104.

Smith M.L., Yang C.J., Metzenberg R.L. and Glass N.L. (1996) Escape from *het-6* incompatibility in *Neurospora crassa* partial diploids involves preferential deletion within the ectopic segment. Genetics **144**: 523-531.

Springer M.L. (1993) Genetic control of fungal differentiation: the three sporulation pathways of *Neurospora crassa*. BioEssays **15(6)**: 365-374.

Staben C. and Yanofsky C. (1990) *Neurospora crassa a* mating-type region. Proc. Natl. Acad. Sci.U.S.A. **87**: 4917-4921.

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

Vellani T.S., Griffiths A.J.F. and Glass N.L. (1994) New mutations that suppress mating-type vegetative incompatibility in *Neurospora crassa*. Genome **37**: 249-255.

Vogel H.J. (1964) Distribution of lysine pathways among fungi; evolutionary implications. Am. Nat. **98**: 435-446.

Vollmer S.J. and Yanofsky C. (1986) Efficient cloning of genes from *Neurospora crassa*. Proc. Natl. Acad. Sci. U.S.A. **83**: 4869-4873.

Watters M.K., Randall T.A, Margolin B.S., Selker E.U. and Stadler D.R. (1999) Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in Neurospora. Genetics **135**: 705-714.

Westergaard M. and Mitchell H.K. (1947) *Neurospora V*. A synthetic medium favoring sexual reproduction. Am. J. Bot. **344**: 573-577.

Wilson J.F. and Dempsey J.A. (1999) A hyphal fusion mutant in *Neurospora crassa*. Fungal Genet. Newsl. **46:** 31.

Wu J. and Glass N.L. (2001) Identification of specificity determinants and the generation of alleles with novel specificity at the *het-c* heterokaryon incompatibility locus of *Neurospora crassa*. Mol. Cell. Biol. **21(4):** 1045-1057.

Wu J., Saupe S.J. and Glass N.L. (1998) Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. Proc. Natl. Acad. Sci. U.S.A. **95:** 12398-12403.

Zhu H., Nowrousian M., Kupfer D., Colot H.V., Berrocal-Tito G., Lai H., Bell-Pedersen D., Roe B.A., Loros J.J. and Dunlap J.C. (2001) Analysis of Expressed Sequence Tags From Two Starvation, Time-of-Day-Specific Libraries of *Neurospora crassa* Reveals Novel Clock-Controlled Genes. Genetics **157:** 1057-1065.

7. APPENDIX

Sequences for the *N. crassa* cDNA clones identified in the yeast two-hybrid screen with TOL (**Tables 1** and **5**; sections **3.1.2** and **3a.3.1**):

7A7

GGCAGCTCCACGACGCTGGAAACAGATGGAGGGGCGGGGAGGATTGTCACCGCGGCGGCGTTGCTTC
ACAAGGCGCGCAACTTTGACGTGCAGACATGGGTGTACAATATCAAGGGACTGCCACCGGATGACGA
CCTGGAAGCGCGGGTAA

7A7p
AAPRRWKQMEGRGGLSPRRRCFTRRATLTCRHGCTISRDCHRMTTWKRG (49 aa)

7C10
GACACCTTCAAGCACACATCAGTCCTGGCAACAACGATGGTCGAGTCCCATCAAGACACGGACATCAA
GATGTCCAACACATCTTCGTCCTCTCAAGGCACTACCGGCCTTGCCATTCGTAACCGGTCCGCAACCCA
AGACATGGCCGCGACCAACACCAGCTCCTCGAGCGTTGCTGTCAAGGACATCATCTTGCATGCTGCCG
CCAATGTTGGCACCGAAATGA

7C10p
TPSSTHQSWQQRWSSPIKTRTSRCPTHLRPLKALPALPFVTGPQPKTWPRPTPAPRALLSRTSSCMLPPMLAP
K (74 aa)

7D3
GAGCAATTCAACAAGGGTACCCAGCAGATCGAAGGTCTCTACACCACTGGCAGCACCCACGGTGTCAC
CCAGGAGTTCAACAAGGGCGGCCAAAACCAGCAGCTCTACAACGGTCAGCAGGTGGGTACTACTGAC
CAGCTGTACAAGACCGGCCAGAACGGTATCCTCCTGA

7D3p (56 aa)
SNSTRVPSRSKVSTPLAAPTVSPRSSTRAAKTSSSTTVSRWVLLTSCTRPARTVSS (56 aa)

7A5
CCGGGATTGA

7A5p (2 aa)
RD (2aa)

7A10 (from the adapter to the stop codon)

CCAAGACGGGGATGGTGACATCGACATGGAAGATTTCGAAATACCCGCTCTTGCATCCACGCCCATGC
CTAGTCGTACTTCATTCCCTCTACCGTTGAGGCCTCTTCAAGAGTCTTCCACTCCCTTGGAAAACTTCAT
ACAGTACCAGCCTCCTGCTGGCAAGCACCAGGCCTCTGTACCAGCACAGACACCAACTAGTCCCGGAC
AAGCTATGAACAGAAAGGACGTGCCCGTGTTTACTGCCACCTCAGCCATTCCCTTTTCGCCCCAGAAG
ACGGAGCCATTGCAGCAGTCTCCGGCGCCGTCTACGCCTGCTTCATTGCCAAGTGTGATCTTTGCTGGA
CAGAAGCAGGAAGAGAAGGAAGAGGTTGTGGAGACGCCAGAGAGGAATTGA

7A10p (second frame, due to construct design with the DNA activation domain vector)

QDGDGDIDMEDFEIPALASTPMPSRTSFPLPLRPLQESSTPLENFIQYQPPAGKHQASVPAQTPTSPGQAMNR
KDVPVFTATSAIPFSPQKTEPLQQSPAPSTPASLPSVIFAGQKQEEKEEVVETPERN (130 aa)

7D8 (*ncvip1*)

ACAACAGCAAACATTTCGGTTCGACCAGTCCTCGACATCTATCACCAAACCGCACACAGATACCATCT
CAACCACAGCAATCATGTCTACAGTCTACGTCAAGAACATCGGCGCCAACACTGAGGAGAAGGACAT
TCGTGCCTTCTTCAGCTTCTGCGGCAAGATCAGCTCCCTCGATGTCACCACCGAAGGCGAGACCAAGT
CCGCCACGGTGACCTTTGAAAAGGAGAGCGCCGCTCGCACTGCTCTTCTCCTCGATCACACCAAGCTT
GGCGAGCACGAGCTCTCTGTCACCTCTGCTTCGGGAGAGCACGCCGATTCCGGCGACAATGTCCACCC
CAAGTCCGACGCCGACCGCGACACAGATGAGATCACACAGGAGGAGAAGCCCCGCGCGCGCGTCCTT
GCCGAGTACTTGGCCAGCGGCTACCTCGTTGCCGACAGCGGCCTGAAGACGGCCATTGCTCTCGACGA
GAAGCACGGCGTCTCTCAGCGCTTCCTGTCGACAATCCAGAACCTCGACCAAAAGTACCACGCAACCG
ACCGCGCCAAGACGGCCGACCAGAGCTACGGCATCACTGCGCGGGCTAACTCCCTGTTCAGCGGCCTC
TCCTCGTACTTTGAGAAGGCCTTGGAGGCGCCTGGCGCCAAGAAGATTGTCGACTTCTACACCACCGG
CTCCAAGCAGGTGCAGGACATCCACAACGAGGCCAAGCGCCTGGCTGAGCTGAAGAAGCAGGAGGCC
GGCGGCAGCTCGTACAAGGCTGCTGGTCTGGACAAGATCTTCGGCGCCGAGAAGGCGCCCGGCCAGG
AGAGCAAGCCCAATGACCAGGTGCCCGGCGCTGCCCCTTCGGATGCTGCCGCCACCGAGTCTAACCAG
CAACCGATTTCCGAGGGAGCGTACCCTGGAACTGCCGAGAAGATTCCCCAGTAA

NCVIP1 (283 aa)

MSTVYVKNIGANTEEKDIRAFFSFCGKISSLDVTTEGETKSATVTFEKESAARTALLLDHTKLGEHELSVTS
ASGEHADSGDNVHPKSDADRDTDEITQEEKPRARVLAEYLASGYLVADSGLKTAIALDEKHGVSQRFLSTI
QNLDQKYHATDRAKTADQSYGITARANSLFSGLSSYFEKALEAPGAKKIVDFYTTGSKQVQDIHNEAKRL
AELKKQEAGGSSYKAAGLDKIFGAEKAPGQESKPNDQVPGAAPSDAAATESNQQPISEGAYPGTAEKIPQ