NON-SELF RECOGNITION IN FILAMENTOUS FUNGI
- THE het-c MEDIATED VEGETATIVE INCOMPATIBILITY
IN NEUROSPORA CRASSA

by

JENNIFER DONGLAN WU
B. Sc., Peking University, 1988
M. Sc., Peking University, 1991

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Biotechnology Laboratory and
Department of Botany

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March, 2000

© Jennifer Donglan Wu, 2000
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 University of British Columbia
Vancouver, Canada

Date March 21, 2000
Abstract

In *Neurospora crassa*, at least eleven *het* loci including the mating-type locus have been identified that regulate non-self recognition during vegetative growth. One of the genetically and molecularly characterized loci, *het-c*, was shown to encode three alleles *het-c^{OR}* , *het-c^{PA}* and *het-c^{GR}*. The three *het-c* alleles encode similar polypeptides with a 34 - 48 amino acid polymorphic region that controls allelic specificity. In an effort to understand the biological role of the *het-c* locus in filamentous fungi, the sequences of *het-c* specificity region from species within *Neurospora* and its related genera were analyzed. The *het-c* locus exhibited a trans-specific mode of evolution (an allele from one species is more closely related to an allele from another species or genus than to other alleles from the same species) based on a phylogenetic analysis of DNA sequences and an increased frequency and number of non-synonymous nucleotide substitutions in the polymorphic domain. This study suggested that the *het-c* locus is under balancing selection for the function of mediating non-self recognition. The polymorphic region in the peptides encoded by the three *het-c* alleles was dissimilar by both amino acid sequences and the pattern of deletion or insertion. To differentiate whether the composition of amino acid sequences or the pattern of insertion/deletion is the critical determinant for *het-c* allelic specificity, chimeric constructs of naturally occurring and artificially generated *het-c* alleles were introduced into *N. crassa* strains with alternative *het-c* specificities. Incompatibility of the transformants was assayed by occurrence of dead hyphal compartments, growth rate and colony morphology. This study suggested that spatial characteristics as affected by the pattern and size of the deletion/insertion within the specificity domain were the primary determinant for *het-c* allelic specificity. Immunoprecipitation studies indicated that non-self recognition is mediated by HET-c heteromeric complex formation during vegetative incompatibility, and that differences in the specificity domain affect the capacity to make heterocomplexes versus homocomplexes.
# Table of Contents

Abstract.................................................................................................................ii

List of Figures........................................................................................................vii

List of Tables...........................................................................................................ix

Acknowledgements.................................................................................................x

Dedication................................................................................................................xi

Chapter 1 General Introduction.............................................................................1

1.1 Self and non-self recognition............................................................................1
1.1.1 Self and non-self recognition in mammals....................................................2
  1.1.1.1 Sexual recognition...................................................................................2
  1.1.1.2 Somatic incompatibility – the MHC complex..........................................2
1.1.2 Sexual incompatibility in plants – the S locus...............................................5
1.1.3 Self and non-self recognition in ascomycete fungi.......................................8
  1.1.3.1 Mating-type and mating recognition in ascomycetes.............................9
  1.1.3.2 Vegetative incompatibility in filamentous ascomycetes..........................10

1.2 Molecular aspects of mating recognition in ascomycetes..............................13
1.2.1 Mating recognition in Saccharomyces cerevisiae.........................................13
1.2.2 Mating recognition in filamentous ascomycetes.........................................16
  1.2.2.1 Life cycle of filamentous ascomycetes....................................................16
  1.2.2.2 The mating-type locus...........................................................................18
  1.2.2.3 Pheromone response pathway..............................................................18
  1.2.2.4 High-Mobility-Group (HMG) and α1 domain proteins.........................19

1.3 Vegetative incompatibility in filamentous ascomycetes.................................20
1.3.1 Vegetative incompatibility in Podospora anserina......................................20
1.3.2 Vegetative incompatibility in N. crassa.......................................................24
  1.3.2.1 Neurospora crassa, the model organism...............................................25
  1.3.2.2 Cellular response to vegetative incompatibility....................................26
  1.3.2.3 Mating-type mediated vegetative incompatibility.................................28

1.4 het-c mediated vegetative incompatibility.......................................................30

1.5 Thesis objectives................................................................................................33
Chapter 2  Evolutionary analyses of the het-c locus

2.1 Introduction

2.2 Materials and Methods

2.2.1 Strains and media

2.2.2 Genomic DNA isolation and Southern analysis

2.2.3 PCR amplification of het-c specificity region

2.2.4 Subcloning of PCR products

2.2.5 Sequencing of PCR products

2.2.6 Sequence analysis

2.2.7 Phylogenetic tree reconstruction

2.2.8 Calculation of synonymous and nonsynonymous nucleotide substitutions

2.2.9 Statistical analysis

2.3 Results

2.3.1 Genomic Southern blot analysis

2.3.2 PCR products and DNA sequences

2.3.3 Amino acid sequences of het-c specificity region

2.3.4 Trans-specific polymorphism of het-c

2.3.5 Number of synonymous and nonsynonymous substitution per site in het-c specificity region

2.4 Discussion

2.4.1 Trans-specific polymorphisms at het-c

2.4.2 het-c allelic lineages were generated prior to the divergence of modern species

2.4.3 het-c allelic lineages

2.4.4 Possible mechanisms of balancing selection at the het-c locus

2.4.5 An excess of nonsynonymous nucleotide substitutions in the het-c polymorphic region

2.4.6 het-c shares evolutionary features with MHC and S loci

Chapter 3  Molecular basis of het-c allelic specificity

3.1 Introduction

3.2 Materials and Methods

3.2.1 Strains and media

3.2.2 Recombinant DNA techniques

3.2.3 Isolation of naturally occurring het-c alleles

3.2.4 Generation of artificial het-c alleles

3.2.5 Generation of vector cassette for construction of chimeric alleles
### 4.3 Results
- 4.3.1 Fusion gene constructs ........................................ 174
- 4.3.2 *het-c* activity of fusion genes .................................. 176
- 4.3.5 Heteromeric complex formation of HET-c proteins ......... 177

### 4.4 Discussion
- 4.4.1 Non-self recognition is mediated via protein heteromeric complex formation ........................................ 185
- 4.4.2 Possible mechanisms for heteromeric complex formation .................................................. 189

### Chapter 5 Concluding Remarks
- 5.1 Summary ........................................................................ 196
- 5.2 General model of how *het-c* mediates vegetative incompatibility ................................................ 197
- 5.3 Future directions .......................................................... 200

### 6. Bibliography ................................................................ 202

### 7. Appendices .................................................................. 221
List of Figures

Figure 1-1. Barrage zone between vegetative incompatible fungal strains ........................................... 11
Figure 1-2. The mating process in *Saccharomyces cerevisiae* ............................................................ 15
Figure 1-3. The life cycle of *N. crassa* ............................................................................................... 17
Figure 1-4. Evan’s blue staining of dead hyphal compartments .............................................................. 27
Figure 1-5. Ultrastructural characteristics of wild type and incompatible heterokaryons ...................... 28
Figure 1-6. Morphology of compatible and incompatible heterokaryons .............................................. 30
Figure 1-7. Features of deduced HET-c peptides from DNA sequences .............................................. 32
Figure 1-8. Amino acid sequence comparison of the variable region II in HET-c .................................... 32
Figure 2-1. Examples of trans-specific polymorphism ............................................................................. 38
Figure 2-2. Schematic diagram illustrating the effect of balancing selection ........................................ 43
Figure 2-3. Diagram of the primer locations for PCR amplification of *het-c* specificity domain .......... 48
Figure 2-4. Southern hybridization of genomic DNA from different species and genera with *het-c*OR probe from *N. crassa* ...................................................................................... 54
Figure 2-5. Inferred amino acid sequences of PCR amplified *het-c* allelic specificity motif from 40 isolates ................................................................................................................. 58
Figure 2-6. Unrooted neighbor-joining tree from the DNA sequence alignment in the *het-c* specificity region including deletion/insertion sites .......................................................... 64
Figure 2-7. Unrooted neighbor-joining tree from the DNA sequence alignment in the *het-c* specificity region excluding deletion/insertion sites ............................................................ 65
Figure 2-8. Unrooted maximum parsimony tree from the DNA sequence alignment in the *het-c* specificity region including deletion/insertion sites ..................................................... 66
Figure 2-9. Unrooted maximum parsimony tree from the DNA sequence alignment in the *het-c* specificity region excluding deletion/insertion sites ..................................................... 67
Figure 2-10. Unrooted maximum likelihood tree from the DNA sequence alignment in the *het-c* specificity region including deletion/insertion sites ....................................................... 68
Figure 2-11. Unrooted maximum likelihood tree from the DNA sequence alignment in the *het-c* specificity region excluding deletion/insertion sites ....................................................... 69
Figure 2-12. Unrooted neighbor-joining tree of *het-c*PA – type alleles .............................................. 70
Figure 2-13. Unrooted neighbor-joining tree of *het-c*OR – type alleles ............................................... 71
Figure 2-14. Unrooted neighbor-joining tree of *het-c*OR – type alleles ............................................... 72
Figure 2-15. Frequency distribution of nonsynonymous and synonymous nucleotide substitution along the codon position of *het-c* specificity domain ..................................................... 75
Figure 3-1. Amino acid sequences of the specificity region of naturally occurring *het-c* alleles .......... 90
Figure 3-2. PCR generation of recombinant *het-c* alleles .................................................................... 94
Figure 3-3. PCR generation of *het-c* alleles by deletion ..................................................................... 95
Figure 3-4. Conserved restriction sites in *het-s* used for chimeric construction ................................. 99
Figure 3-5. Principals for chimeric construction and transformation assay for *het-c* specificity .......... 101
Figure 3-6. Plasmid map of the vector *pCB1004* .................................................................................. 104
Figure 3-7. Plasmid map of the vector pOKE103
Figure 3-8. Colony morphology of transformants containing the wild-type het-c alleles into het-c^OR, het-c^PA and het-c^GR background
Figure 3-9. Average growth rate of het-c transformants
Figure 3-10. Hyphae morphology and occurrence of compartmentation of the three classes of het-c transformants
Figure 3-11. Representative morphology of transformants of naturally occurring mutant het-c^OR chimeric constructs
Figure 3-12. Hyphal compartmentation and death of Ndl3923 (p26) chimeric construct
Figure 3-13. Amino acid sequences comparison of HET-c^PA, P26 and P26m specificity region
Figure 3-14. Phenotype comparison of het-c^PA, p26, p26m transformants after three days of vegetative growth on Vogel’s selective medium
Figure 3-15. Average growth rate comparison of het-c^PA, p26, p26m transformants
Figure 3-16. Amino acid sequence comparison of HET-c^PA, HET-c^OR, HET-c^GR, PD1, PD2, PD3 and PD4 specificity region
Figure 3-17. Average growth rate comparison of PA, GR, pd1, pd2, pd3 and pd4 transformants of het-c^OR, het-c^PA and het-c^GR background
Figure 3-18. Phenotypes of vegetative growth of allele pd1, pd2, pd3 and pd4 transformants
Figure 3-19. Hyphae morphology and occurrence of dead hyphal compartments in transformants when pd1, pd2, pd3 and pd4 chimeric constructs were introduced into het-c^OR, het-c^PA and het-c^GR spheroplasts
Figure 3-20. Amino acid sequence comparison of PO1 and PO2
Figure 3-21. Phenotype of transformants of alleles pol and po2 chimeric constructs
Figure 3-22. Average growth rate of alleles pol, po2 chimeric construct transformant
Figure 3-23. Hyphal compartmentation and death by the introduction of pol, po2, chimeric constructs into het-c^PA and het-c^OR strains
Figure 3-24. Sequence comparison of PO3, PO4, OD1 and GO1
Figure 3-25. Sequence comparison of DEL3, HET-c^GR, HET-c^OR and HET-c^PA
Figure 3-26. Phenotype of transformants of allele del3
Figure 3-27. Average growth rate of allele del3 transformants
Figure 3-28. Sequence comparison of P26 and PD1
Figure 3-29. Sequence comparison of PD3 and OR
Figure 3-30. Secondary structure prediction for HET-c specificity region
Figure 3-31. Proposed protein conformation in the variable domain for determining HET-c specificity
Figure 3-32. Comparison of the predicted secondary structure of PD1 and HET-c^OR in the specificity region
Figure 4-1. Outline of het-c^OR:GFP fusion construction
Figure 4-2. Outline of het-c^OR:HA and het-c^OR(PA):HA fusion gene construction
Figure 4-3. Outline of het-c^OR:HA fusion gene construction in pCB1004 vector
Figure 4-4. Summary of predicted fusion gene products
Figure 4-5. Western blot analysis of mouse anti-GFP immunoprecipitation samples
  with mouse anti-GFP antibody
Figure 4-6. Western blot analysis of mouse anti-GFP immunoprecipitation samples
  with polyclonal anti-HA antibody
Figure 4-7. Western blot analysis of polyclonal anti-HA immunoprecipitation samples
  with anti-HA antibody
Figure 4-8. Western blot analysis of polyclonal anti-HA immunoprecipitation samples
  with monoclonal anti-GFP antibody
Figure 4-9. The putative leucine rich haptad repeat region of HET-c polypeptides
Figure 4-10. Schematic model for stable HET-c heterocomplex formation
Figure 5-1. General model of how het-c mediates vegetative incompatibility
List of Tables

Table 2-1. Strains used for phylogenetic analysis.........................................................45
Table 2-2. Comparison of synonymous and nonsynonymous nucleotide substitutions
per 100 sites in het-c polymorphic and non-polymorphic region...............................76
Table 3-1. List of strains as transformation recipients.......................................................88
Table 3-2. List of strains used for heterokaryon incompatibility tests.............................89
Table 3-3. Oligonucleotide primers for the construction of het-c alleles..........................96
Table 3-4. Specific primer sets and template for the 1st round PCR to generate specific
alleles..........................................................................................................................97
Table 3-5. Phenotypes of the three classes of het-c transformants......................................110
Table 3-6. Summary of the phenotypes of the introduction naturally occurring
het-c<sup>OR</sup> - type chimeric constructs into N. crassa..............................................115
Table 3-7. Average growth rate of the transformants containing naturally occurring
het-c<sup>OR</sup> - type chimeric constructs.........................................................................116
Table 3-8. Transformation of naturally occurring het-c<sup>GR</sup> - type chimric
constructs into N. crassa..............................................................................................118
Table 3-9. Average growth rate of transformants containing naturally occurring het-c<sup>GR</sup>
chimric constructs..........................................................................................................119
Table 3-10. Transformantion of naturally occurring het-c<sup>PA</sup> - type alleles
into N. crassa................................................................................................................122
Table 3-11. Average rate of growth of transformants containing naturally occurring
het-c<sup>PA</sup> chimeric constructs .....................................................................................123
Table 3-12. Summary of the phenotypes of the transformants of het-c alleles
pd1, pd2, pd3 and pd4........................................................................................................129
Table 3-13. Summary of phenotype of pol1 and po2 transformants .................................134
Table 3-14. Summary of phenotype of po3, po4, od1 and gol transformants....................139
Table 3-15. Phenotypes of co-transformants of pol/po2, pd1, pd2 and pd3.........................145
Table 4-1. Summary of the phenotypes of introducing fusion gene constructs into
N. crassa..........................................................................................................................177
Acknowledgements

I am very grateful to:

Dr. Louise Glass for giving me the opportunity to complete my study in her laboratory and for her supervision, precious advice and support during the course of this work.

The members of my thesis committee, Dr. Jim Kronstad, Dr. Sally Otto and Dr. Michel Roberge for their advice on the research and the dissertation.

In particular, Dr. Anthony Griffiths for providing me the space in his laboratory for the completion of my research after the departure of Dr. Louise Glass; Dr. Jenneat Whitton for generously providing the phylogenetic analysis program and the computer.

Dr. Carl Douglas, for the understanding, support and advice on my career choice.

Rostam Namdari (Monka) for being there all the time, listening to all my frustrations as well as joys (Dollars!), for his caring, encouragement, and long-time friendship.

Dorian Medd who came to my life with love at such a subtle phase of my life.

My family in China, for their love, their patience with and understanding of my never-ending student life.

All the colleagues present and past in the Glass Lab and the Griffiths Lab, for their help, advice and many more.

Friends from Botany Department and Biotechnology Laboratory. Especially people from Tony Glass’s lab, Kronstad’s lab, Finlay’s lab, Jeffrey’s lab.

My friends for sharing my frustrations and JOYS with, Tania Thenu, Suman, Greg Lezee, Reza, Fatima, Fiona, Rebacca, ......

Dr. Mary Berbee and Dr. Daniel Haydon for their help with phylogenetic analysis.

Dr. Ken-ichiro Ishida for suggestions about the phylogenetic analysis and critical reading of this part of the dissertation.

Botany Department and Biotechnology Laboratory for the friendly environment and financial support.
Dedications

Dedicated to my forever-beloved parents, for their patience with and understanding of my never-ending student “life”.

Also to my dear siblings, in particular to my sister Rose (Yueji) and her family, for their support and taking my obligations towards our senior parents.

Also, in the memory of my dear Mother.

___ Donglan
Chapter 1  General Introduction

1.1  SELF AND NON-SELF RECOGNITION

Self and non-self recognition is a fundamental biological process that is essential for sexual reproduction, defense against pathogen invasion and maintenance of individuality and integrity for an organism. Two apparently diametrically opposed mechanisms: self-incompatibility or non-self incompatibility, i.e., caused by identical or different gene products, may induce the recognition and subsequent discrimination reactions between cells or organisms. The self-incompatibility mechanism is active in the sexual reproduction process of some hermaphroditic species, favoring outcrossing within the species (Esser and Blaich, 1973). The non-self incompatibility mechanism is active in somatic recognition between cells of different individuals, favoring the integrity of the individual or species. The paradox of recognition of identity and difference applies to a variety of systems. In 1968, Burnet first suggested that the ability to distinguish “self” and “non-self” appeared initially in predatory protozoa, in which the amoeba has to discriminate between engulfed prey and itself (Burnet, 1968). Cell-cell recognition was also suggested to be associated with the emergence of sexuality (Monroy and Rosati, 1979).

In prokaryotes, cell-cell recognition is best elucidated in the so-called “sexual events” in the bacteria Escherichia coli. The conjugation events between F+ and F− bacterial strains are mediated by recognition of the cell surface glycoprotein encoded by the F “sex” plasmid (Willetts and Skurray, 1980). In eukaryotic cells, self and non-self recognition occurs during both sexual and asexual (somatic) growth, although different mechanisms are involved.
Identical gene products may cause self-incompatibility during sexual reproduction and different gene products may cause non-self incompatibility in somatic growth. These parallel phenomena are discussed in mammals, plants and ascomycete fungi.

1.1.1 Self and non-self recognition in mammals

1.1.1.1 Sexual recognition

In mammals, the male and female individuals have specialized sex organs for production of different reproductive cells (gametes). Reproduction requires the initial recognition of the female and male reproductive cells. The female reproductive cell, the egg, is surrounded by a glycoprotein layer that is responsible for mediating recognition of the male reproductive cell and thereafter fertilization (Beleil and Wassarman, 1980; Hanada and Chang, 1972). Genetically, however, sexual incompatibility does not restrict inbreeding. The male and female reproductive cells may carry identical genetic information (except for the male determination chromosome) in a random mating population. Promotion of outbreeding is generally achieved by dioecism (for review, see Esser and Blaich, 1973).

1.1.1.2 Somatic incompatibility

In mammals, the major histocompatibility complex encoded by an array of Mhc loci controls somatic non-self recognition and subsequent activation of defense mechanisms (for reviews, see Bjorkman, 1990; Klein, 1980; 1986). MHC molecules were initially defined as the main target antigens in organ transplantation reactions. When organ grafts are exchanged between adult individuals, either of the same species (allografts) or of different species (xenografts), they were usually rejected in response to the foreign antigens on the surface of
the grafted cells. These genetically "foreign antigens" or cell surface proteins were identified as major histocompatibility complex molecules (MHC molecules).

The major histocompatibility complex (MHC) is a multigene family encoding cell-surface glycoproteins that play a key role in immune response (Klein, 1986). The MHC glycoproteins and the T cell surface receptors (TcRs) are the key elements of specificity in the T cell response to foreign antigens. The MHC glycoproteins can be considered as antigen-presenting molecules. The antigenic peptide must be bound by a MHC glycoprotein to form a complex that can be recognized by the TcRs.

The MHC encodes two categories of cell surface glycoproteins called class I and class II molecules, both of which are transmembrane heterodimers (reviewed in Bjorkman, 1990; Hughes and Yeager, 1998). Class I MHC molecules are expressed on the surface of all somatic cells and function to destroy virus-infected cells by presenting foreign-processed peptides to TcRs of cytotoxic T cells. The class II molecules are expressed on the surface of cells such as B cells, dendritic cells, macrophages, and epithelial cells. The class II molecules bind foreign peptides (foreign antigens) to form a ligand complex and present the complex to the TcRs of helper T cells to activate an immune response. The class II molecule is a heterodimer made of an α chain and a β chain, both of which are encoded in the class II region of the MHC complex.

The class I MHC molecules are non-covalently linked heterodimers, consisting of an α chain (or heavy chain) and a molecule called β2-microglobulin (β2m). The α chain is made up of three extracellular domains (designated as α1, α2 and α3), a transmembrane region and a cytoplasmic domain. The β2m molecule only has a single domain extracellularly. The α chains are encoded within the MHC complex by the class I loci, of which there are three in
humans, HLA-A, HLA-B, and HLA-C. The class I loci are polymorphic. The \( \beta_2 m \) molecule is encoded by a non-polymorphic locus outside the MHC complex.

Both classes of MHC play an essential role in immune defense against foreign pathogens. However, there appears to be no specific mechanisms for distinguishing self-proteins from foreign antigens at the level of peptides binding to MHC molecules (Lorentz and Allen, 1988; 1989). The discrimination between self-proteins and non-self proteins occurs at the level of the TcRs that can only recognize foreign (non-self) peptides in the context of self-MHC molecules. When the cell is infected with a virus or other intracellular antigens, the viral peptides or other foreign peptides will compete with endogenous self-peptides for binding to the MHC molecules on the surface of the infected cell. Only when the TcRs of the T cells encounter the complex of self-MHC molecules and foreign peptides, an aggressive T cell response will be stimulated.

Both class I and class II MHC molecules appear to have similar peptide binding regions (PBR) in their extracellular domains (Bjorkman et al., 1987a, b; Brown et al., 1988, 1993). Highly polymorphic features were found in the PBR region of both molecules that appear to enhance the capacity of each molecule to bind various foreign peptides (Klein, 1986; Falk et al., 1991).

MHC molecules are expressed on the cells of all higher vertebrates. They were first found in mice and called H-2 antigens (histocompatibility-2 antigens, Klein, 1979). In humans they are called HLA antigens (human-leukocyte-associated antigens) because they were first demonstrated on leukocytes (for review, see Klein, 1986). They have been identified in birds (Bourlet et al., 1988), amphibians (Flajnik et al., 1991) and fish (Hashimoto et al., 1990; Miller and Withler, 1996). Each individual has five or more loci
encoding MHC molecules. Many of the functional *MHC* loci are the most polymorphic loci known in higher vertebrates (Klein, 1979, 1986; Klein et al., 1993; Ono et al., 1993; Miller and Withler, 1996); that is, within a species there is a large number of alleles at each locus and each allele is present at a relatively high frequency in the population. For example, more than 149 alleles have been identified at the human *MHC* class I locus *HLA-B* (Klein, 1986; Parhem and Ohta, 1996); more than 126 *DRB* alleles have been identified at the human *MHC* class *II DRB* locus (Yuhki and O'Brien, 1997).

Despite the enormous diversity of MHC molecules, polymorphism and allelic lineages at the functional class I and class II *MHC* loci are very stable, persistent and shared between different species. These evolutionary characteristics of *MHC* loci are believed to be maintained by balancing selection (Doherty and Zinkernagel, 1975; Hughes and Nei, 1988, 1989). Supporting evidence for this conclusion includes: (1) a large number of alleles occur at functional *MHC* loci (reviewed in Klein, 1986); (2) allele frequencies are distributed more or less evenly in a population (Hedrick and Thomason, 1983); (3) Trans-specific polymorphism was found at both class I and class II loci (i.e. certain alleles of one species are generally more similar to certain alleles of another species than they are to other alleles of the same species (Klein, 1980; Figueroa et al., 1988; Lawlor et al., 1988; Mayer et al., 1988; McConnell et al., 1988)); (4) the rate of nonsynonymous substitutions per site at the variable PBR is higher than that of synonymous substitutions per site (Hughes and Nei, 1988; 1989).

### 1.1.2 Sexual incompatibility in plants – the *S* locus

In many flowering plant species, sexual reproduction is mediated by the gametophytic or sporophytic self-incompatibility locus, the *S* locus (for reviews, see Nasrallah and
Nasrallah, 1986; Haring et al., 1990; Charlesworth, 1994; 1995). Self-incompatibility is a mechanism by which many flowering plants prevent self-fertilization and promote outbreeding. By contrast to the MHC complex, it is a mechanism for self-recognition that results in rejection of self-pollen by the female somatic tissues. In certain species, the specificity of the self-incompatibility interaction has been shown to be controlled by the S locus (for review, see Charlesworth, 1994; 1995). The two self-incompatibility systems differ in the phenotype expressed by the pollen grain during pollination. In gametophytic self-incompatibility, represented in the Solanaceae (Nicotiana, Petunia, and Solanum species), the single S-allele of the haploid pollen grain determines self-incompatibility (reviewed by Kao and McCubbin, 1996). Fertilization is prevented if the S allele carried by the pollen is identical to one of the S alleles of the recipient plant. In the sporophytic system, represented in the Brassicaceae and exemplified by the genus Brassica, the interaction of the two S alleles that are borne by the two diploid parent plants (sporophyte) determines the self-incompatibility phenotype (Goodwillie, 1997). The cellular mechanism of recognition of both types of self-incompatibility is not well understood (for reviews, see Clarke and Newbigin, 1993; Kao and McCubbin, 1996; Nasrallah, 1997). The S loci exhibit allelic polymorphism at high levels in both systems (reviewed by Charlesworth, 1995).

The S locus of the gametophytic self-incompatibility system encodes a pistil-specific protein, the S protein, which functions in the recognition and rejection of pollen bearing the same allele (Haring et al., 1990). In members of the Solanaceae, the S locus encodes glycoproteins with ribonuclease (RNase) activity (McClure et al., 1989). The role of RNase activity in self-incompatibility is unclear although several models have been proposed (Kao and McCubbin, 1996; Hiscock et al., 1998).
Sporophytic self-incompatibility systems have been described in a number of species and extensively studied in Brassicaceae. In *Brassica* species that exhibit self-incompatibility, two tightly linked genes, *SLG* and *SRK*, at the *S* locus are involved in the recognition reaction with self and non-self pollen (for review, see Trick and Heizmann, 1993; Hatakeyama et al., 1998). The *SLG* (*S* locus glycoprotein) gene encodes a highly polymorphic secreted glycoprotein that localizes primarily to the cell wall of epidermal cells of the stigma (Nasrallah et al., 1985, 1987; for review, see Hiscock et al., 1998), whereas the *SRK* (*S* locus receptor kinase) gene encodes a transmembrane serine/threonine protein kinase with an extracellular domain (*S* domain) that spans the plasma membrane (Stein et al., 1996). Several lines of evidence suggest that both of these genes are required for the recognition reaction of self-incompatibility (see reviews, Nasrallah and Nasrallah, 1986; Hiscock et al., 1998; Nasrallah, 1997). Current models predict a central role for SRK in the initiation of the self-rejection response, yet the molecular and biochemical basis of the signaling pathway is unknown (Hiscock et al., 1998).

The gametophytic incompatibility system has been most extensively studied in the Solanaceae. *S* alleles have also been sequenced from species in the Papaveraceae, Onagraceae and Gramineaceae (for reviews, see Charlesworth, 1995; Kao and McCubbin, 1996). The *S* alleles of the sporophytic incompatibility system have also been identified in species of other families (the Primulaceae and the Asteraceae) besides those in the Brassicaceae (reviewed by Charlesworth, 1995). A high degree of DNA sequence variability was found among *S* alleles, even between alleles from the same species. As with the *MHC*, *S* loci exhibit several features of loci under balancing selection: 1) A large number of alleles occur in a population. In the Solanaceae, populations commonly harbor as many as 30 to 50 alleles (Lane and Lawrence,
1993; Sakamoto et al., 1998). 2) $S$ allele polymorphisms are shared among species and arose prior to speciation (Ioerger et al., 1990). That is, alleles at the $S$ locus in the Solanaceae show “trans-specific” mode of evolution (Ioerger, et al., 1990; Richman et al., 1995). 3) Allele polymorphisms have persisted for an evolutionarily significant period of time (Hinata et al., 1995). Polymorphisms observed in the $S$ alleles in members of the Solanaceae have persisted for at least 36 million years (Clarke and Kao, 1991; Richman et al., 1996). 4) An excess of nonsynonymous nucleotide substitutions over synonymous nucleotide substitutions was observed at shared polymorphic sites (Clark and Kao, 1991; Hinata et al., 1995). A high nonsynonymous nucleotide substitution rate was observed in the polymorphic region of both $SLG$ and $SRK$ loci (Hinata et al., 1995) and at the shared polymorphic sites among $S$ alleles in the Solanaceae (Clarke and Kao, 1991).

1.1.3 Self and non-self recognition in ascomycete fungi

In fungi, non-self recognition events occur in both sexual and vegetative reproduction stages. Sexual reproduction in outcrossing species is governed by mating-type genes that determine the pattern of mating among individual strains. Recognition mediated by these genes allows sexual development. The mating recognition system mostly restricts inbreeding and promotes outbreeding and thereby maintains genetic variability within a population. Filamentous fungi also possess a somatic recognition system to restrict hyphal fusion between genetically non-identical individuals. Such a system is referred as vegetative or heterokaryon incompatibility and is genetically regulated by $het$ (for heterokaryon) loci.
In filamentous ascomycetes, mating recognition and vegetative incompatibility in general do not interfere with each other. However, in some species, such as *Neurospora crassa* or *Ascobolus immersus*, the mating type genes also function as *het* genes that regulate vegetative incompatibility during the asexual reproductive phase.

1.1.3.1 Mating-type and mating recognition in ascomycetes

In fungi, the mating-type locus (*mat*) controls mating recognition. The mating type system in fungi was first identified in the zygomycete *Rhizopus* (Blakeslee, 1904), in which successful mating only occurs between two sexually compatible strains. Isolates of this fungus are morphologically indistinguishable, but sexually divided into two compatible mating groups designated as “+” and “-”. Subsequently, mating types were discovered in ascomycetes (Edgerton, 1914) and basidiomycetes (Bauch, 1927).

Several mating strategies exist in ascomycetes, including heterothallism, homothallism and pseudohomothallism (reviewed in Nelson, 1996). Heterothallic species (such as *N. crassa, N. sitophila, N. discreta*, and *Cochliobolus heterostrophus*) are self-sterile. Mating requires strains of opposite mating-types to come together during the sexual process. Homothallic species (such as *N. pannonica, N. dodgei*) are self-fertile, in which a single haploid nucleus carries all the information necessary for the sexual process. Pseudohomothalllic species superficially resemble homothallic species in that they are also self-fertile. Unlike homothallic species, pseudohomothalllic species (such as *N. tetrasperma, Podospora anserina*) have two opposite mating-type nuclei that are compartmentalized into a single ascospore. Thus, pseudohomothallic species do not exhibit mating-type associated
Vegetative incompatibility; stable heterokaryons are formed by nuclei carrying opposite mating-type genes.

The mating system in heterothallic and pseudohomothallic ascomycetes is bipolar - a one-locus, two-"allele" mating system (for reviews, see Glass and Kuldau, 1992). Mating recognition is mediated at two steps: one involves recognition that controls initial plasmogamy; the other involves recognition of opposite mating-type nuclei for post-fertilization development. In the unicellular ascomycetes, studies on mating recognition have been conducted in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (for review, see Herskowitz, 1989). In filamentous ascomycetes, studies on mating recognition have been conducted mainly in *N. crassa*, *P. anserina*, and *C. heterostrophus* (for reviews, see Glass and Kuldau, 1992; Coppin et al., 1997; Kronstad and Staben, 1997). The molecular mechanisms of mating recognition in representative species will be described in Section 1.2.1 and 1.2.2.

1.1.3.2 Vegetative incompatibility in filamentous ascomycetes

In filamentous fungi, somatic cell-cell recognition regulates the success of hyphal fusion and heterokaryon formation during vegetative growth phase. When vegetative compatibility permits heterokaryon formation, heterokaryotic growth occurs. In many fungal species, however, vegetative incompatibility is observed by the occurrence of a barrage between incompatible mycelia, which is a macroscopic line of numerous degenerated and dead hyphal compartments that underwent anastomoses (Beguret et al., 1994; Esser and Blaich, 1994; Figure 1-1). In the laboratory, vegetative incompatibility can be detected directly by forced heterokaryon formation between two auxotrophic strains. Heterokaryons
formed between two strains that have alternative alleles at a \textit{het} locus are inhibited in their growth, displaying hyphal compartmentation and subsequent destruction by a lytic process (for reviews, see Glass and Kulda, 1992; Leslie, 1993; and Begueret et al, 1994).

\textbf{Figure 1-1.} Barrage zone between fungal strains that are vegetatively incompatible. \textbf{a,} Barrage formation between vegetatively incompatible \textit{P. anserina} isolates. A contact between two incompatible strains is shown by the solid black arrow. The other arrow shows a contact between compatible strains (Adapted from Begueret et al., 1994). \textbf{b,} Cross section of a log colonized by different isolates of basidiomycetes (Adapted from Esser and Blaich, 1994). Border region between incompatible mycelial is shown by dark lines that result from regeneration of apical hyphae after the death of incompatible hyphal compartments.

Understanding vegetative incompatibility and its genetic control mainly comes from studies in the filamentous ascomycetes \textit{N. crassa} and \textit{P. anserina}, both of which are in the Sordariaceae (for reviews, see Glass and Kulda, 1992; Leslie, 1993; Begueret et al., 1994). In a given species, numerous \textit{het} loci occur, e. g. at least ten \textit{het} loci have been identified in \textit{N. crassa} and nine have been identified in \textit{P. anserina} (for reviews, see Begueret et al.,
Allelic differences at any of these *het* loci between two individuals are sufficient to cause a vegetative incompatibility reaction upon hyphal anastomosis. Molecular and cellular aspects of vegetative incompatibility in these two model species will be discussed in detail in Section 1.3.

Vegetative incompatibility is widespread in fungi. Its biological significance in fungal populations is however, still unclear. It is not known whether vegetative incompatibility and vertebrate graft rejection phenomena arose by convergent evolution (Esser and Blaich, 1973) in each group or whether vegetative incompatibility is a primitive trait of allore cognition or immune response (Lane, 1981). Vegetative incompatibility is believed to prevent transmission of conventional parasites such as mycoviruses and detrimental cytoplasmic elements, which are spread via hyphal fusion (Caten, 1972; Anagnostakis, 1983; Debets et al., 1994; Nauta and Hoekstra, 1994; van Diepeningen et al., 1997). Advantages such as preventing nuclear invasion or genomic replacement after anastomosis has also been suggested (Todd and Rayner, 1980; Rayner et al., 1984; Rizzo and Harrington, 1992; Rizzo and May, 1994; for review, see Worrall, 1997). Vegetative incompatibility may also limit outbreeding and genetic polymorphism within the species, especially in pseudohomothallic species. In particular, it limits the chance of somatic genetic recombination that is important in imperfect fungi (e.g. *Aspergillus niger* and *Fusarium oxysporum*) that have lost their sexuality (Caten, 1971).
1.2 MOLECULAR ASPECTS OF NON-SELF RECOGNITION IN ASCOMYCETES

Understanding the molecular mechanisms of mating recognition in fungi has been explored with a combination of genetic and biochemical studies (for reviews, see Kahmann and Bolker, 1996; Kronstad and Staben, 1997). The molecular and biochemical basis of vegetative incompatibility is not as well characterized. Nonetheless, the understanding of mating recognition may provide concepts for how vegetative incompatibility is accomplished.

In fungi, non-self recognition in mating is accomplished through two distinct molecular mechanisms. One is via a pheromone signaling pathway to choose a mating partner. The other involves heterodimerization of proteins to gain new regulatory functions favoring nuclear recognition and proliferation after mating. These mechanisms have precedents in both yeast and filamentous ascomycete mating systems.

1.2.1 Mating recognition in Yeast

The best characterized mating recognition system in fungi is that of the yeast *S. cerevisiae*. Mating recognition in this organism relies on a complex signal transduction pathway via mating-type specific pheromone receptors (for review, see Johnson, 1995). Cells of one mating-type respond only to pheromone produced by cells of the opposite mating-type.

The laboratory maintained heterothallic budding yeast *S. cerevisiae* has two mating-types, *a* and *α*, encoded by the *MAT* locus. Three cell types with different properties exist in its life cycle: the haploid *a* cell and α cell and the diploid *a/α* cell. *S. cerevisiae* does not have
specialized reproductive structures and is competent for mating during vegetative growth without the requirement for nutrient depletion. Mating occurs only between haploid $\alpha$ cells and $\alpha$ cells; $\alpha/\alpha$ cells cannot mate (for review see Borkovich, 1996). The $\text{mat } \alpha$ locus encodes MAT$\alpha$1 and MAT$\alpha$2; the $\text{mat } a$ locus encodes MAT$a$1 (for review, see Kues and Casselton, 1992). In haploid cells, MAT$\alpha$1 activates $\alpha$-cell specific genes, i.e. pheromones and pheromone receptors; MAT$\alpha$2 represses the expression of $a$-specific haploid genes. The MAT$a$1 is not required for the expression of any $a$-specific haploid genes (Fields, 1990; for review, see Kues and Casselton, 1992). The haploid $\alpha$ cells and $\alpha$ cells secrete cell-specific peptide pheromones, the $\alpha$-pheromone and the $\alpha$-pheromone, respectively. Each haploid cell possesses a seven transmembrane pheromone receptor on the cell membrane for binding pheromones produced by the opposite mating type cells (Burkholder and Hartwell, 1985; Hagen et al., 1986; for review, see Cross, 1988). For instance, the pheromone receptor Ste2 protein on the membrane of the $\alpha$ cells binds $\alpha$ pheromone, while the Ste3 protein on the surface of the $\alpha$ cells is the receptor for $a$ pheromone. The pheromone-receptor ligand induces expression of a set of genes in the $\alpha$ cells and $\alpha$ cells via a G-protein and MAP kinase signaling pathway, resulting in cell differentiation and formation of a protrusion, or "schmoo". Such a structure allows physical contact and thus conjugation and nuclear fusion between the $\alpha$ cell and $\alpha$ cell to form a diploid $\alpha/\alpha$ cell (Figure 1-2). In diploid $\alpha/\alpha$ cells, the two homeodomain proteins MAT$a$1 and MAT$\alpha$2 form an $a1-\alpha2$ heterodimer that acts as a transcriptional repressor of haploid specific genes such as RME1 (repression of meiosis) and allows meiosis and sporulation to ensue under appropriate nutritional conditions (Dranginis, 1990; Covitz et al., 1991).
**Figure 1-2.** The mating process in *Saccharomyces cerevisiae*. The events that occur during conjugation are shown. Cells are shown as starting the mating process in the G1 phase of the cell cycle before spindle-pole body duplication. G1 arrest mediated by the presence of a- and α-pheromones will result in the arrest of cells at this point prior to mating. (Adapted from Cross et al., 1988).
1.2.2 Mating recognition in filamentous ascomycetes

Mating recognition in filamentous ascomycetes has features in common with yeast via mating type-specific pheromone response signaling pathways controlled by mating-type loci (for review, see Kronstad and Staben, 1997). In the following sections, an overview will focus on the well-characterized mating systems from the two representative filamentous ascomycetes, *N. crassa* and *P. anserina*.

1.2.2.1 Life cycle of filamentous ascomycetes

The principal events of the life cycle are similar between *N. crassa* and *P. anserina* except that asci in the pseudohomothallic species, *P. anserina*, contain four ascospores in which two nuclei of opposite mating-types are enclosed. As an example of filamentous ascomycetes, the life cycle of *N. crassa* is presented in Figure 1-3. It consists of vegetative and sexual growth phases. Asexual (or vegetative) reproduction generally occurs by either mycelial propagation or by the production of numerous vegetative spores, known as conidia (Alexopoulos, 1962). Under nutrient starvation, particularly nitrogen depletion (Westergaard and Mitchell, 1947), *N. crassa* switches from vegetative growth to sexual development. The sexual cycle begins with the differentiation of female reproductive structures, protoperithecia, which are composed of a spherical mass of tightly coiled hyphae. A specialized hypha, called trichogyne, grows directionally from the protoperithecium towards a male cell (typically hyphae or asexual spores) of the opposite mating-type and fuses with the male fertilizing cell. This directional growth is believed to be mediated by mating-type (*A* and *a*) pheromones (Bistis, 1981; 1983). Nuclei from the male are transported into the ascogonium after cell
fusion. The proliferation of haploid nuclei of both mating-types results in the formation of dikaryotic ascogenous hyphae, followed subsequently by nuclear fusion between opposite mating-type nuclei. The diploid nucleus immediately undergoes meiosis and eight haploid ascospores are formed in a sac-like structure called an ascus.

**Figure 1-3.** Life cycle of *N. crassa*. (Adopted from Nelson, 1996).
1.2.2.2 The mating-type locus

The mating-type locus in filamentous ascomycetes contains dissimilar DNA sequences that do not seem to have an allelic relationship, for which the termed idiomorph was adopted (Metzenberg and Glass, 1990). Idiomorphic structure is common to most known fungal mating-type genes, including those from the unicellular ascomycetes *S. cerevisiae* and *S. pombe*.

Mating-type idiomorphs of *N. crassa* are termed *mat A* and *mat a* (Metzenberg and Glass, 1990). The *mat A* idiomorph contains three mating-type genes, *mat A-1*, *mat A-2* and *mat A-3*, among which only *mat A-1* is essential for mating recognition (Ferreira et al., 1996). The *mat a* idiomorph contains only one transcription unit *mat a-1*, which is essential for mating recognition (Chang and Staben, 1994). The mating idiomorphs of *P. anserina* are similar to those of *N. crassa*, but are designated as *mat+* and *mat-* (for review, see Glass and Kuldau, 1992). The *mat+* idiomorph has a single gene, *FPR1*, that is homologous to *mat a-1* in *N. crassa*. The *mat-* idiomorph contains three genes, *FMR1*, *SMR1* and *SMR2*, with similarity to *mat A-1*, *mat A-2* and *mat A-3*, respectively (Debuchy and Coppin, 1992; Debuchy et al., 1993). Two genes, *FPR1* and *FMR1*, mediate mating recognition.

1.2.2.3 Pheromone response pathway

Similar to *S. cerevisiae*, it is believed that mating-type in filamentous ascomycetes functions to control recognition mechanisms leading to fertilization through a pheromone response pathway. In *N. crassa*, males of each mating-type secrete a pheromone to which females of the opposite mating-type respond by orienting growth of their trichogynes (Bistis, 1983). Mating-type mutants of *N. crassa* strains do not orient hyphal growth towards
pheromone source or produce pheromone (Bistis, 1996), suggesting that mating-type specific pheromones and their receptors are likely to be regulated by mating-type gene products.

1.2.2.4 High-Mobility-Group (HMG) and α1 domain proteins

A DNA binding motif, the high-mobility-group (HMG) domain was found in the MAT proteins of *N. crassa* (MAT a-1 and MAT A-3), and *P. anserina* (FPR1 and SMR2) (reviewed by Coppin et al., 1997). A subregion of the MAT A-1 (*N. crassa*) and FMR1 (*P. anserina*) polypeptide has similarity to the protein encoded by the *mata1* gene of *S. cerevisiae*. It has been proposed that MAT a-1 in *N. crassa* controls mating via the DNA binding activity of its HMG domain (Philley and Staben, 1994). Whether MAT a-1 and MAT A-1 dimerize, as in the case of *S. cerevisiae* α2-a1, to confer a novel cell-type is unknown. In *P. anserina*, the two genes *FPR1* (homolog of *mat a-1* in *N. crassa*) and *FMR1* (homolog of *mat A-1* in *N. crassa*) are responsible for mating specificity. Yeast two-hybrid data failed to support an interaction of FPR1 and FMR1, suggesting that some other mechanisms may exist for the regulation of post-fertilization functions (Coppin et al., 1997). However, yeast two-hybrid analysis showed an interaction between MAT A-1 and MAT a-1 in *N. crassa* (C. Staben, personal communication).
1.3 VEGETATIVE INCOMPATIBILITY IN FILAMENTOUS ASCOMYCETES

Vegetative incompatibility has been described in many filamentous ascomycetes (reviewed by Esser and Blaich, 1994; Leslie, 1993; Begueret et al., 1994). The genetic basis of incompatibility has been elucidated in several filamentous ascomycetes such as *Aspergillus nidulans* (Grindle, 1963), *Cryphonectria parasitica* (Anagnostakis, 1977), *N. crassa* (Garnjobst, 1953, 1955; Perkins, 1988) and *P. anserina* (Bernet, 1967). However, studies conducted in most species are limited to population analysis due to the lack of genetic and molecular tools. Understanding the cellular, biochemical and molecular basis of vegetative incompatibility has mainly been conducted in two species of filamentous ascomycetes, the pseudohomothallic species, *P. anserina*, and the heterothallic species, *N. crassa*. Genes involved in vegetative incompatibility have been cloned and molecularly characterized from both species (for reviews, see Begueret et al., 1994; Glass and Kuldau, 1992; Saupe et al., 1996; Saupe and Glass, 1997). However, many aspects of the cellular and biochemical bases of vegetative incompatibility remain unclear.

1.3.1 Vegetative incompatibility in *Podospora anserina*

In *P. anserina*, genes present at *het* loci can be involved in both allelic and non-allelic interactions (for review see Begueret et al., 1994). Nine *het* genes have been identified. Three *het* loci, *het-s* (Turcq et al., 1990), *het-c* (Saupe et al, 1994) and *het-e* (Saupe et al., 1995) have been molecularly characterized. The *het-c* locus of *P. anserina* has no similarity with the *het-c* locus in *N. crassa* that will be described in section 1.3.2. Among the nine known *het* loci, *het-c, het-d, het-e, het-r* and *het-v* are involved in non-allelic incompatibility; non-allelic
genetic interactions between \textit{het-c/het-e}, \textit{het-c/het-d}, and \textit{het-r/het-v} have been characterized (Bernet, 1967; 1992).

The \textit{het-s} locus mediates allelic incompatibility and encodes two mutually incompatible alleles, \textit{het-s} and \textit{het-S}. Hyphal fusion between wild type \textit{P. anserina} \textit{het-s} and \textit{het-S} strains triggers a vegetative incompatibility response and cell death (for review, see Esser and Blaich, 1994). Disruption of the \textit{het-s} locus does not affect cell viability and fertility (Turcq et al., 1991). The \textit{het-s} and \textit{het-S} alleles encode similar peptides of 289 amino acids, differing by 14 amino acids. Chimeric construction and site-specific mutagenesis showed that a single amino acid difference between the two polypeptides was sufficient to elicit vegetative incompatibility (Deleu et al., 1993). How this difference in amino acid composition mediates vegetative incompatibility is unclear.

A neutral allele, \textit{het-s*}, was also identified in wild type isolates (Bernet, 1967). Strains containing the \textit{het-s*} allele are compatible with both \textit{het-s} and \textit{het-S} strains. A \textit{het-s*} strain can switch to \textit{het-s} specificity spontaneously at a low frequency or after fusion with the hyphae of a \textit{het-s} strain (Deleu et al., 1993). Expression of the HET-s protein was found at the same level in \textit{het-s} and \textit{het-s*} strains, which led to the conclusion that a post-translational modification of the HET-s protein is necessary for the phenotypic differences between \textit{het-s} and \textit{het-s*} strains (Begueret et al., 1997). Several lines of evidence indicate that the HET-s protein shares features with the prion protein (Begueret et al., 1997; personal communication). Firstly, both are able to form homodimers. Secondly, like the prion protein, the HET-s can exist in two forms which differ in the sensitivity to proteolytic enzymes. And thirdly, overexpression of HET-s can enhance the switch of \textit{het-s*} strain to \textit{het-s} strains.
The het-c and het-e loci are multiallelic and are involved in non-allelic incompatibility. Genes have been cloned from these two loci. The het-e locus contains an open reading frame encoding a polypeptide of 1356 amino acids with two interesting sequence motifs. The predicted HET-E protein contains functional domains similar to those present in α and β subunits of heterotrimeric G proteins: the GTP-binding motif at the N-terminal and Gβ homologous domain (or the WD40 repeated domain) at the C-terminal (Saupe et al., 1995a). Mutational analysis and in vitro binding assays indicated that both the GTP binding domain and the WD40 repeats were required for vegetative incompatibility. It was suggested that HET-E might be involved in signal transduction mediating vegetative incompatibility (Espagne et al., 1997). Four alleles were identified at the het-c locus in wild-type strains. The four het-c alleles encode similar polypeptides of 208 amino acids with similarity to glycolipid transfer proteins (Saupe et al., 1994; 1995b). Polypeptides encoded by the different het-c alleles contain 16 polymorphic positions and a single amino acid difference was found sufficient to modify allele specificity (Saupe et al, 1995b). It was suggested that selection pressure maintained the high degree of polymorphism at the het-c locus based on the observation of excess nonsynonymous nucleotide substitutions at polymorphic sites among the four het-c alleles (Saupe et al, 1995b). Disruption of het-c locus dramatically affects ascospore formation, suggesting that het-c is essential in the life cycle of P. anserina. However, how interaction between HET-E and HET-C mediates vegetative incompatibility is unclear.

Studies on biochemical mechanisms mediating vegetative incompatibility in P. anserina are under way by the isolation of genes (idi) induced upon vegetative incompatibility and suppressor mutations (or modifier mutations, mod mutations). Several idi
(idi-1, 2, 3) and mod (mod-Al, mod-D and mod-E) genes have been molecularly characterized. The putative IDI proteins are small proteins with signal peptides and were proposed to be involved in regulating different steps of the incompatibility reaction (Bourges et al., 1998). Most of the mod mutations affect both protoperithecial development and vegetative incompatibility, suggesting that the incompatibility reaction in P. anserina may be related to the regulation of differentiation. It has been proposed that het genes may control some differentiation steps in the life cycle of P. anserina, such as the transition from the quiescent to the proliferation state or the formation of female organs (Bernet, 1992).

The mod-Al mutation has no effect on allelic vegetative incompatibility interactions but suppresses incompatibility resulting from all three non-allelic systems (Bernet et al., 1973). The mod-Al mutation restores the growth of incompatible strains but does not suppress the cell lytic reaction (Bernet et al., 1973). The putative MOD-A polypeptide contains a proline-rich domain with similarity to SH3-binding motifs (Barreau et al., 1998). It was suggested that MOD-A is involved in differentiation and functions as key regulator of growth arrest associated with vegetative incompatibility (Barreau et al., 1998). The suppressor mutation mod-Bl does not modify the phenotype of incompatibility of the non-allelic system, but the presence of both mod-Al and mod-Bl mutations suppresses the growth arrest and cell lytic reactions of non-allelic systems. The mod-C1 mutation specifically suppresses incompatibility in the het-R/het-V strains. The predicted MOD-D polypeptide is a Gα subunit of a heterotrimeric G protein that is involved in the cAMP signal transduction pathway. The mod-D mutant strains present developmental defects and partial suppression of vegetative incompatibility (Loubradou et al., 1999). A gene coding for an adenylate cyclase, PaAC, was also identified as a partial suppressor of a mod-D mutant allele (Loubradou et al., 1999).
The mod-E mutation suppresses both vegetative incompatibility and developmental defects due to the mod-D mutation. The putative MOD-E is a member of the HSP90 family, a possible component of a signaling pathway involved in both vegetative incompatibility and development (Loubradou et al., 1997).

Genetic and mutational studies suggested that the vegetative incompatibility response was triggered by the co-expression of incompatible het genes (for review, see Begueret et al., 1994). Based on these studies, a “poison heteromeric complex” model of het gene products has been proposed for mediating vegetative incompatibility in P. anserina (Begueret et al., 1994). It was proposed that the heteromeric complex formed by incompatible het gene products might have direct lethal effects on the cell, perhaps by disrupting gene regulation related to P. anserina development (Bourges et al., 1998). The molecular characterization of modifier genes that mediate incompatible interactions and are involved in differentiation may help explain the link between incompatibility and developmental regulation. However, no direct evidence is available that supports the hypothesis of protein heteromeric complex formation during vegetative incompatibility.

1.3.2 Vegetative incompatibility in N. crassa

Studies of natural populations of N. crassa found that no two isolates from the same geographical area were able to form a stable heterokaryon (Mylyk, 1976). Genetic determinants that limit the formation of stable heterokaryons have been well studied in N. crassa (Mylyk, 1975; Perkins, 1975). At least ten het loci, including the mating type locus have been identified that regulate the capacity to form stable heterokaryons between different isolates (Mylyk, 1975; Perkins, 1975; 1988). The mating-type locus differs from other het
loci because it also regulates entry into sexual reproduction (Beadle and Coonradt, 1944; Sansome, 1946). Morphological features and cellular responses associated with vegetative incompatibility have been described (Perkins, 1988; Jacobson et al., 1998). Two het loci, mat and het-c, have been genetically and molecularly characterized. The het-6 locus was cloned recently (Smith et al., in press). Unlike P. anserina, non-allelic interactions between het loci have not been described in N. crassa, although they may occur at het-6 (Smith et al., in press).

1.3.2.1 Neurospora crassa, the model organism

Neurospora is blessed with well-known advantages for genetic research. It is haploid with a relatively small genome. It has a short generation time and a fast growth rate. It grows on simple, defined media and can be preserved. Large numbers of sexually produced progeny can be recovered for mapping of genes. Among Neurospora species, N. crassa was chosen as a model to study vegetative incompatibility for several reasons. All ten het loci have been well mapped genetically (Perkins, 1988). N. crassa translocation strains are available that allow the construction of partial diploids of particular het loci that may be either heterozygous or homozygous (Perkins, 1975). There are a large number of laboratory-maintained isogenic strains available for heterokaryon tests. Gene mutations can be easily achieved by a unique process termed RIP (Repeated Induced Point mutation) that allows directional mutagenesis to specific DNA sequences (Selker et al., 1987; Cambareri et al., 1989).
1.3.2.2 Cellular response to vegetative incompatibility

The cellular and biochemical mechanisms behind vegetative incompatibility remain largely unknown although morphological features have been described. One characteristic of vegetative incompatibility is hyphal compartmentation and lytic death (Figure 1-4). A description of hyphal compartmentation and death during vegetative incompatibility has been provided by recent ultrastructural studies of incompatible partial diploids (Jacobson et al., 1998). No apparent differences were found between hyphal compartmentation and death resulting from incompatible het-c, het-6, or mat partial diploids or heterokaryons (Jacobson et al., 1998; Glass, personal communication), suggesting that each het gene may initiate a common hyphal death pathway (Jacobson et al., 1998). Ultrastructural changes were obvious in incompatible partial diploid cells and transformed cells as compared with wild-type cells, i.e. hyphal compartmentation by septal plug formation, shrinkage of cell membrane, degradation of cell organelles, cytoplasm vacuolization and loss of nuclear contents. Incompatible hyphae commonly show an “invasion” event in the death process; dying hyphae were “invaded” by adjacent healthy hyphae, which may subsequently undergo hyphal compartmentation and death (Figure 1-5).

Ultrastructural features of incompatible hyphae exhibit similarities with programmed cell death (PCD) in animal and plant cells (Jacobson et al., 1998; Fraser and Evan, 1996; Havel and Durzan, 1996). Programmed cell death can be defined as an orderly process where the dying cell actively participates in its own death (Columbano, 1995; Fraser and Evan, 1996). It is in contrast to necrosis or degenerative death caused by the external environment (Columbano, 1995). Programmed cell death is morphologically characterized by “blebbing” of the plasma membrane, formation of apoptotic bodies and chromatin condensation, and
biochemically characterized by fragmentation of nuclear DNA at internucleosomal sites (Cohen et al., 1994). Nuclear DNA degradation was detected by the TUNEL (Terminal UTP Nick End Labeling) assay in het-c incompatible transformants, suggesting that vegetative incompatibility might have some biochemical similarities to programmed cell death (Marek et al., 1999). However, whether the hyphal death triggered by vegetative incompatibility shares common genetic pathways with programmed cell death in animals or plants remains to be determined.

Figure 1-4. Hyphal compartments in het-c incompatible heterokaryons. The dead and dying hyphae were stained within the vital stain, Evan’s blue (arrows).
1.3.2.3 Mating-type associated vegetative incompatibility in *N. crassa*

The mating-type locus has two functions; it controls not only mating but also vegetative incompatibility. Fusion of reproductive structures from strains of opposite mating-types (mat A and mat a) initiates sexual reproduction. Hyphal fusion between mat A and mat a strains elicits an incompatibility reaction that results in the death of fused hyphal compartments (Beadle and Coonradt, 1944; Garnjobst, 1953; Gross, 1952; Staben and Yanofsky, 1990); only mat A-1 and mat a-1 are required for vegetative incompatibility (Glass et al., 1990; Staben and Yanofsky, 1990; Ferreira et al., 1996).
The *mat A*-*l* gene encodes a 293 amino acid polypeptide with a region of homology to the MATα1 mating-type polypeptide of *S. cerevisiae* (Glass et al., 1990). Functional dissection of the MAT A-1 polypeptide showed that a region of 99 amino acids at the N-terminal was required for female fertility and that a region of 111 amino acids at the N-terminal (including the above 99 amino acids) was sufficient for conferring vegetative incompatibility, while male fertility required amino acids from position 1 to 227 (Saupe et al., 1996a).

The *mat a* encodes a 382 amino acid polypeptide containing specific regions essential for mating and vegetative incompatibility. Functional dissection of the MAT a-1 polypeptide showed that the HMG domain was required for DNA binding and mating but dispensable for vegetative incompatibility (Philley and Staben, 1994). In contrast, the C-terminal acidic region was found necessary for mediating vegetative incompatibility but non-essential for mating. It has been proposed that the *mat a* gene controls vegetative incompatibility via interaction of the C-terminal part with another factor, while the HMG domain controls mating via its DNA binding activity (Philley and Staben, 1994).

Mating-type associated vegetative incompatibility must be somehow suppressed during sexual reproduction to allow the proliferation of opposite mating-type nuclei in ascogenous hyphae. A recessive mutation unlinked to the mating-type locus, *tol* (for tolerant) was found to suppress mating-type associated vegetative incompatibility without affecting mating ability (Newmeyer, 1970). The *tol* locus encodes a 1011 amino acid polypeptide containing an amphipathic α-helix and a leucine-rich repeat domain (Shiu and Glass, 1999). How MAT A-1, MAT a-1 and TOL mediate mating-type associated incompatibility during vegetative growth is unclear.
1.4 *HET-C MEDIATED VEGETATIVE INCOMPATIBILITY IN N. CRASSA*

The *het-c* locus was originally identified in forced heterokaryons (Garnjobst, 1953; Garnjobst and Wilson, 1956) and then further characterized by using translocation strains generating partial duplications of the *het-c* locus (Perkins, 1975). Incompatible heterokaryons or partial diploids heterozygous for *het-c* display a slow-growing, curly, flat, aconidiating phenotype (Perkins, 1988; Figure 1-6).

![Figure 1-6. Morphology of compatible and incompatible *het-c* heterokaryons.](image)

Figure 1-6. Morphology of compatible and incompatible *het-c* heterokaryons. a, compatible *het-c* heterokaryons formed between two auxotrophic *het-c*\(^A\) strains. b. Incompatible heterokaryons formed between two auxotrophic strains that are *het-c*\(^{PA}\) and *het-c*\(^{OR}\) at the *het-c* locus, respectively.

The *het-c* locus is located on the left arm of LGII, ~ 1 map unit centromere distal to *pyr-4*. The *het-c* locus encodes three distinct and mutually incompatible alleles, *het-c*\(^{OR}\), *het-c*\(^{PA}\) and *het-c*\(^{GR}\) (Saupe and Glass, 1997). The *het-c*\(^{OR}\) allele encodes a 966 amino acid...
polypeptide, with a signal peptide in the N-terminal, an amphipathic alpha-helical region and a glycine-rich domain at the carboxy-terminal (Figure 1-7). Introduction of a mutated het-c^{OR} allele with deletion of sequence coding for the entire glycine rich region or the amphipathic alpha-helical region into a het-c^{PA} strain gave compatible transformants, suggesting that both the glycine rich region and the amphipathic alpha-helical region are required for mediating vegetative incompatibility (Saupe et al., 1996b). Introduction of a het-c^{OR} allele with deletion of the sequence coding for the signal peptide into a het-c^{PA} strain gave incompatible transformants (S. J. Saupe and R. Todd, personal communication). The disruption of het-c^{OR} does not affect the vegetative growth or fertility of N. crassa, but strains containing an inactivated het-c^{OR} allele display dual compatibility with both het-c^{OR} and het-c^{PA} strains (Saupe et al., 1996b)

The three het-c alleles encode similar peptides, which share approximately 85% amino acid sequence identity. Sequence comparison revealed three variable regions between the three HET-c polypeptides. The first variable region, from amino acid positions 149-212 in HET-c^{OR}, has 78% amino acid sequence identity between HET-c^{OR}, HET-c^{PA} and HET-c^{GR}. The second variable region (region II, Figure 1-7), from positions 247-284 in HET-c^{OR}, showed only 45% sequence identity between the three polypeptides. Within this variable region, the HET-c polypeptides differed by 20 amino acid substitutions and three deletion or insertion events (Figure 1-8). The third variable region, from positions 487-610 in HET-c^{OR}, shares 96% sequence identity between HET-c^{OR} and HET-c^{PA}, while the identity between HET-c^{OR} or HET-c^{PA} with HET-c^{GR} was only 68%. Chimeric construction and transformation assays demonstrated that the variable region II (position 247 to 284 in HET-c^{OR}) controls het-c allelic specificity (Saupe and Glass, 1997)
Figure 1-7. Features of deduced HET-c peptides from the DNA sequences. The HET-c\textsuperscript{OR} is a putative peptide of 966 amino acids, including a signal peptide at the N-terminal, an amphipathic alpha helical (coiled-coil) motif and a glycine rich domain at the C-terminal (Saupe et al., 1996b). The deduced three peptides HET-c\textsuperscript{OR}, HET-c\textsuperscript{PA} and HET-c\textsuperscript{GR} share more than 80% amino acid identity and differ in three variable domains, I, II and III (Saupe and Glass, 1997). Chimeric construction suggested that the variable domain II (amino acid 247-284 in HET-c\textsuperscript{OR}) confers \textit{het-c} allelic specificity (Saupe and Glass, 1997).

Figure 1-8. Amino acid sequence comparison of the variable region II in HET-c. Note that the three polypeptides differ by several amino acids and the pattern of insertions/deletions. HET-c\textsuperscript{OR} has a putative insertion of five amino acids in length; HET-c\textsuperscript{PA} has a putative insertion of 15 amino acids in length. Asterisks indicate identical sites.
1.5 THESIS OBJECTIVES

Vegetative incompatibility is a universal phenomenon among filamentous fungi. However, the biological significance of vegetative incompatibility in fungal populations is unclear. It has been proposed that vegetative incompatibility may be a fungal "immune" system. The question whether vegetative incompatibility is a bona fide non-self recognition system and whether it confers a selective advantage for non-self recognition, such as the MHC complex in vertebrates and the S system in plants, is unclear. In N. crassa, the het-c locus has been molecularly well characterized, and polymorphisms at the het-c locus have been observed in a number of N. crassa isolates (Saupe and Glass, 1997). The link between het-c polymorphisms and allelic specificity allowed me to examine the evolutionary features of the het-c locus and thus shed light on the biological significance of vegetative incompatibility.

Morphological aspects associated with vegetative incompatibility have been extensively documented in N. crassa (Perkins, 1988; Jacobson et al., 1998). However, the molecular mechanisms behind vegetative incompatibility remain largely unknown. The isolation of naturally occurring het-c alleles in the study of het-c evolutionary features provided tools for the dissection of the molecular basis of allelic specificity. However, an understanding of how amino acid and spatial variability in the specificity domain affects vegetative incompatibility does not address the question of how non-self recognition is mediated. The molecular basis of non-self recognition is well understood in the yeast mating system. Is non-self recognition during vegetative incompatibility mediated via HET-c heteromeric complex formation as the case for a1/a2 in S. cerevisiae?
In the effort to understand the biological significance and the molecular mechanisms of vegetative incompatibility, the specific objectives of my thesis were:

1. An examination of the evolutionary features of the \textit{het-c} locus in the Sordariaceae to define the biological significance of vegetative incompatibility in fungal populations.

2. To delineate the determinants for \textit{het-c} allelic specificity.

3. To determine whether non-self recognition is mediated via the mechanism of HET-c heteromeric complex formation.

In Chapter 2, I analysed the evolutionary feature of the \textit{het-c} locus in the Sordariaceae. The evolutionary features between the \textit{het-c} locus and known loci responsible for non-self recognition, such as the \textit{MHC} and the \textit{S} loci, were compared.

In Chapter 3, specificity of the naturally occurring \textit{het-c} alleles that were isolated in Chapter 2 was assayed to examine whether amino acid substitutions in the variable domain of HET-c primarily affect \textit{het-c} specificity. Since HET-c\textsuperscript{OR}, HET-c\textsuperscript{PA} and HET-c\textsuperscript{GR} in \textit{N. crassa} also differ by deletions/insertions, artificial \textit{het-c} alleles were generated to examine how the deletions/insertions may affect \textit{het-c} specificity. A model of how \textit{het-c} specificity may be determined is proposed.

In Chapter 4, I performed immunoprecipitation assay to determine whether the HET-c proteins form heteromeric complexes during vegetative incompatibility. A possible model of \textit{het-c} non-self recognition is proposed.
Chapter 2  Evolutionary analyses of the *het-c* locus

"This polarity between the Self-Assertive and Integrative tendencies is inherent in the concept of hierarchic order; and a universal characteristic of life." – Arthur Koestler, *The Ghost in the Machine*.

Some sequences in this chapter were contributed by Dr. Sven Saupe (presently at Université de Bordeaux, Bordeaux, France) and also published in Saupe and Glass, 1997. Most of the results described here are published in Wu J., S. J. Saupe and N. L. Glass (1998).

2.1 INTRODUCTION

The phenomenon of vegetative incompatibility is widespread in filamentous fungi, however, its biological significance in natural populations is unclear. Todd and Rayner (1980) suggested that vegetative incompatibility is a protection mechanism of the "fungal individual" against invasion of foreign nuclei and cytoplasm. In a population genetic study, Nauta and Hoekstra (1994) suggested that selection against transmission of a harmful cytoplasmic element is a more plausible explanation for vegetative incompatibility than against a nuclear gene invasion. Vegetative incompatibility has been shown to restrict horizontal transmission of cytoplasmic elements between individuals in several species, such as *N. crassa* (Caten, 1972; Debets et al., 1994), *Cryphonectria parasitica* (Anagnostakis, 1983) and *Aspergillus niger* (van Diepeningen et al., 1997).

It has been argued whether vegetative incompatibility in fungi functions as a non-self recognition system or whether its existence is a simple consequence of variation in genes with critical cellular functions (Begueret et al., 1994). If vegetative incompatibility confers non-self recognition function in fungal populations, selection may favor polymorphisms at *het* loci. Therefore, *het* loci should exhibit hallmarks of loci under balancing selection and
display similarities to the MHC and S loci, such as large number of alleles existing in a population, persistence of trans-specific polymorphisms and enhanced nonsynonymous nucleotide substitutions at polymorphic sites.

In N. crassa, het-c encodes a putative glycine rich polypeptide containing a consensus sequence for a signal peptide. A 34-48 amino acid region of HET-c determines het-c allelic specificity (Saupe and Glass, 1997). The identification of a particular polymorphic domain that confers het-c allelic specificity has allowed us to examine the diversity of het-c allelic types at the het-c locus in different species within the Sordariaceae and to derive inferences about the evolutionary history of selection in the het-c gene.

2.1.1 Trans-specific polymorphism

Trans-specific polymorphism refers to a peculiar pattern in which allelic lineages (or gene polymorphisms) have been passed from an ancestral species to its descendant species (Klein et al., 1993). In phylogenetic analyses, trans-specific polymorphisms are inferred when an allele from one species is found to be more closely related to an allele from another species or genus than to other alleles from the same species; that is, alleles cluster in phylogenetic analyses rather than species. Observations of trans-specific polymorphisms have often been restricted to loci believed to be under balancing selection. Two such systems have been studied extensively, the major histocompatibility complex (MHC) of vertebrates (for reviews, see Klein, 1986; Klein et al., 1998; Hughes and Yeager, 1998) and the self-incompatibility S loci in flowering plants (for reviews, see Klein et al., 1998; Charlesworth and Awadalla, 1998). However, neutral trans-specific polymorphisms may occur under other circumstances, such as a relatively short period since speciation and a large founding
population size. Such examples have been documented in the sequences of 3'UTR or introns of five randomly selected nuclear genes in the Lake Victoria haplochromines (Nagl, 1998).

The mammalian MHC loci are good examples of trans-specific polymorphism (Ayala, et al., 1994). As an example, in phylogenetic analysis of the MHC class II DRB alleles, the human allele HLA-DRB1*0302 clustered with the chimpanzee allele Patr-DRB1*0305. The alternative human allele HLA-DRB1*0701 clustered with the chimpanzee allele Patra-DRB*0702. The human allele HLA-DRB1*0302 differs from the chimpanzee allele Patr-DRB1*0305 in the exon encoding the peptide binding region (PBR) by 13 nucleotide substitutions and from the other human allele HLA-DRB1*0701 by 31 nucleotide substitutions, while the chimpanzee allele Patra-DRB*0702 differs from the human allele DRB*0701 by only two nucleotide substitutions. This relationship is reflected in a phylogenetic tree of the four DRB alleles (Figure 2-1). Phylogenetic analysis of DRB alleles indicated that the DRB allelic lineages arose prior to the separation of the human and chimpanzee lineages and have persisted in subsequent populations since their divergence.

Trans-specific polymorphism at the MHC loci has also been described in other mammals such as Old World primates and apes (Boyson et al., 1995, 1996; Cadavid and Watkins, 1997; for review, see Cadavid et al., 1997) and other vertebrates including birds (Bourlet et al., 1988), amphibians (Flajnik et al., 1991) and fish (Hashimoto et al., 1990; Graser et al., 1996).

The trans-specific nature of the S locus polymorphism has been demonstrated for both the gametophytic (Ioerger et al., 1990; Richman et al., 1996) and the sporophytic (Dwyer et al., 1991) self-incompatibility systems. In a sample of S alleles from the gametophytic self-incompatibility system in the Solanaceae (Figure 2-1, Ioerger et al., 1990), some S alleles of
Nicotiana alata clustered with certain $S$ alleles of Petunia inflata rather than with other $S$ alleles from other $N. alata$ isolates. For example, the $S_f$ allele of $P. inflata$ clustered with $SF_{11}$ allele of $N. alata$; and the $S_3$ allele of $P. inflata$ clustered with the $S_z$ allele of $N. alata$. The $S$ allele clustering pattern suggests that $S$ allele polymorphisms arose prior to the divergence of $P. inflata$ and $N. alata$.

Figure 2-1. Examples of trans-specific polymorphism at the $MHC DRB1$ locus and plant self-incompatibility $S$ locus. The allelic lineages of four alleles at the $MHC DRB1$ locus from humans ($HLA$) and chimpanzees ($Pan troglodgtes, Patr$) predate the divergence of the human and chimpanzee lineages (Adapted from Ayala et al., 1994). Similarly, the allelic lineages of four $S$ alleles predate the divergence of $Nicotiana alata$ and $Petunia inflata$ (modified from Ioerger et al., 1990). Note that many more alleles are present within the populations than are shown here.
2.1.2 Selection systems

Mutations can be selectively neutral, deleterious, or advantageous (reviewed in Klein et al., 1998). An allele with negative selective value (disadvantage) is usually eliminated from the gene pool soon after its appearance. An allele with an initially positive selective value can also be lost, but if it isn’t, it can either become fixed (in the case of directional selection) or established at an immediate equilibrium frequency (in the case of balancing selection, see below). A neutral allele will eventually become eliminated or fixed (i.e., replace all other alleles at that locus) by genetic drift.

Most neutral polymorphisms are not expected to persist in populations for a very long time. Theoretical studies predict that a neutral allele, if not lost, takes an average of $4N_e$ generations to become fixed, where $N_e$ is the effective population size (which is roughly the number of gene pairs at a given locus that are passed on from one generation to the next) (Kimura, 1983; Nei, 1987). Assuming a long-term effective population size of $10^4$ and generation time of 15 years for humans, it will take only 600,000 years for a neutral allele to become fixed. Thus, neutral polymorphism is a relatively transient phenomenon with respect to evolutionary time.

If the allele occurs in the vicinity of a site under balancing selection, neutral alleles may persist in a population for much longer periods of time than expected based on the attributes of the population (i.e., a fast speciation or a oversized population) (Nagl, 1998). Genetic theory predicts that balancing selection slows down or completely hinders the fixation of alleles and thus allows the persistence of allelic polymorphisms over extensive periods of time (Maruyama and Nei, 1981). Under balancing selection, allele polymorphism can persist without fixation for as long as selection pressure persists and the population
remains at a reasonable size (Takahata and Nei, 1990). Two forms of balancing selection have received a great deal of attention in the literature, namely, overdominant selection and frequency-dependent selection. In the case of overdominant selection, heterozygotes have a selective advantage over either homozygote. In the case of frequency-dependent selection, an allele is advantageous when it is rare but becomes disadvantageous when common.

Balancing selection is the most efficient mechanism to maintain polymorphisms and trans-specific allele lineages (Takahata, 1990). A high level of polymorphism in both the MHC and S loci is believed to be maintained by balancing selection (Doherty and Zinkernagel, 1975; Hughes and Nei, 1988; 1989). At the MHC class I DRB locus, trans-specific allele lineages have persisted for at least 29 million years in humans, 36 million years in chimpanzee, 34 million years in gorillas and 46 million years in macaques (Klein, 1986; Mayer et al., 1988; McConnell et al., 1988; Lawlor et al., 1988; Ayala et al., 1994). At the gametophytic self-incompatibility S loci, trans-specific allele lineages have been retained for at least 36 million years (Ioerger et al., 1990). It is unclear whether selection at MHC loci is overdominant (Doherty and Zinkernagel, 1975; Maruyama and Nei, 1981) or frequency-dependent (Takahata and Nei, 1990), or some other form of selection (e. g., maternal-fetal interactions) (Clarke and Kirby, 1966; Ohta, 1991; Hughes and Nei, 1992a). Polymorphism at the S locus is believed to be maintained primarily by frequency-dependent selection (Vekemans and Slatkin, 1994; Schierup et al., 1997).
2.1.3 Synonymous and nonsynonymous nucleotide substitutions

There are two classes of nucleotide substitutions in protein coding regions: synonymous and nonsynonymous. Synonymous (or silent) substitutions are nucleotide substitutions that do not result in amino acid changes, whereas nonsynonymous nucleotide substitutions are those that change amino acids within a protein. The synonymous nucleotide substitution rate largely reflects the mutation rate. Theoretical studies predict that overdominant selection should enhance nonsynonymous nucleotide substitutions (Lopez et al., 1982). An excess of nonsynonymous nucleotide substitutions was observed in both the PBR coding region of MHC molecules (Nei and Gojobori, 1986; Cadavid et al., 1997; Hughes and Nei, 1989) and the polymorphic sites of the S alleles (Hinata et al., 1995; Clark and Kao, 1991).

2.1.4 Hypotheses on the biology of vegetative incompatibility

Natural populations of N. crassa are highly polymorphic for het genes (Mylyk, 1975; Debets et al., 1994; 1998). Two hypotheses have been proposed for the biological significance of vegetative incompatibility in relationship to the polymorphisms of het genes in fungal populations. One proposes that vegetative incompatibility is a bona fide non-self recognition system in fungi. Under this hypothesis, incompatibility systems have been selected to prevent heterokaryon formation between genetically unlike individuals (non-self) to limit the horizontal transfer of infectious elements, such as mycoviruses, debilitated organelles and deleterious plasmids (Caten, 1972; Lane, 1981). The alternative hypothesis is that vegetative incompatibility is a consequence of genetic variation in genes with critical
cellular functions in which heteroallelism (co-existence of two allelic types) is unfavorable and that vegetative incompatibility is not directly favored (reviewed in Begueret et al., 1994).

The two hypotheses imply different evolutionary features of het genes as depicted in Figure 2-2. Under the first hypothesis, selection favors polymorphisms at het loci for non-self recognition to prevent the invasion of unfavorable cytoplasmic elements from unlike individuals. In this case, alleles at het loci should exhibit evolutionary features similar to other loci that regulate non-self recognition, such as MHC and S loci, i.e. long persistence of polymorphisms in allelic lineages. Under the second hypothesis, polymorphisms at het loci could be found transiently within populations due to random mutations but are not maintained by selection over long periods of time. Thus, one would expect monophyletic relationships of het genes within contemporary species.

To distinguish between these two hypotheses, I describe the evolutionary features of the het-c locus in species in the Sordariaceae. Evidence that supports balancing selection acting at the het-c locus includes the persistence of trans-specific polymorphisms and excess nonsynonymous nucleotide substitutions at the polymorphic region in the het-c specificity domain. The study supports the hypothesis that vegetative incompatibility is under balancing selection, which is consistent with its role as a non-self recognition system.
Figure 2-2. Schematic diagram illustrating the effect of balancing selection when alleles are descended from an ancestral allele pool. Patterns of dots do not indicate different alleles, but rather different modern species found carrying alleles. One would expect that alleles within a species are more closely related if alleles are neutral (left); and that functionally related alleles in different species should be more closely related than alleles within a species (trans-specific polymorphism) if balancing selection has been acting (right).
2.2 MATERIALS AND METHODS

2.2.1 Strains and media

For the het-c phylogenetic analysis, we used 40 strains representing eleven species and three different genera within the Sordariaceae (Table 2-1). All strains were obtained from Fungal Genetics Stock Center (FGSC; Department of Microbiology, University of Kansas Medical Center, Kansas City, KS). All strains were maintained on Vogel’s vegetative growth media (Vogel, 1964).

2.2.2 Genomic DNA isolation and Southern analysis

Genomic DNA was prepared as described by Oakley et al. (1987). Two micrograms of genomic DNA was digested with restriction enzymes Pst I and Sac I according to the manufacturer’s instructions. Gel electrophoresis and transfer of DNA to nylon filters (Nytran+, Schleicher & Schuell, Keene, NH) was done according to Sambrook et al. (1989). A 3.9 kb fragment of het-cOR genomic clone was radiolabeled with $^{32}$P using the T7 quick primer kit (Pharmacia, Baie d’Urfe, PO).
<table>
<thead>
<tr>
<th>Strain designation</th>
<th>FGSC number</th>
<th>Origin</th>
<th>Source of sequences</th>
<th>het-c type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Neurospora crassa</em></td>
<td>4711</td>
<td>Haiti</td>
<td>Saupe and Glass, 1997</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>4832</td>
<td>Ivory Coast</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>2489</td>
<td>Louisiana</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>1455</td>
<td>Unknown</td>
<td>Saupe and Glass, 1997</td>
<td>GR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>4709</td>
<td>Haiti</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>430</td>
<td>Ivory Coast</td>
<td>Saupe and Glass, 1997</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>847</td>
<td>Louisiana</td>
<td>Saupe and Glass, 1997</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>1693</td>
<td>Louisiana</td>
<td>Saupe and Glass, 1997</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>1824</td>
<td>Pakistan</td>
<td>Saupe and Glass, 1997</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>4481</td>
<td>Louisiana</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>4486</td>
<td>Louisiana</td>
<td>This study</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>4499</td>
<td>Louisiana</td>
<td>This study</td>
<td>GR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>1945</td>
<td>Florida</td>
<td>Saupe and Glass, 1997</td>
<td>GR</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>967</td>
<td>Liberia</td>
<td>Saupe and Glass, 1997</td>
<td>GR</td>
</tr>
<tr>
<td><em>N. intermedia</em></td>
<td>2316</td>
<td>Florida</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. intermedia</em></td>
<td>1940</td>
<td>Florida</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. intermedia</em></td>
<td>3290</td>
<td>Oahu</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. intermedia</em></td>
<td>3721</td>
<td>Kauai</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. intermedia</em></td>
<td>1799</td>
<td>Malay</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. discreta</em></td>
<td>3228</td>
<td>Texas</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. discreta</em></td>
<td>5923</td>
<td>Texas</td>
<td>S. J. Saupe</td>
<td>X</td>
</tr>
<tr>
<td><em>N. discreta</em></td>
<td>6788</td>
<td>New Guinea</td>
<td>This study</td>
<td>GR</td>
</tr>
<tr>
<td><em>N. discreta</em></td>
<td>6793</td>
<td>Brazil</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. sitophila</em></td>
<td>963</td>
<td>France</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. sitophila</em></td>
<td>5940</td>
<td>Tahiti</td>
<td>S. J. Saupe</td>
<td>OR</td>
</tr>
<tr>
<td><em>N. sitophila</em></td>
<td>5941</td>
<td>Tahiti</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. tetrasperma</em></td>
<td>6538</td>
<td>Tahiti</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. tetrasperma</em></td>
<td>1270</td>
<td>Unknown</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. pannonica</em></td>
<td>7221</td>
<td>Hungary</td>
<td>S. J. Saupe</td>
<td>PA</td>
</tr>
<tr>
<td><em>N. dodgi</em></td>
<td>1692</td>
<td>Puerto Rico</td>
<td>S. J. Saupe</td>
<td>GR</td>
</tr>
<tr>
<td><em>Gelasinospora sp.</em></td>
<td>8241</td>
<td>Yucatan Penn.</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>Gelasinospora sp.</em></td>
<td>8239</td>
<td>Yucatan Penn.</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>Gelasinospora sp.</em></td>
<td>8243</td>
<td>Yucatan Penn.</td>
<td>This study</td>
<td>PA</td>
</tr>
<tr>
<td><em>Gelasinospora sp.</em></td>
<td>8242</td>
<td>Yucatan Penn.</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>Sordaria brevicollis</em></td>
<td>7140</td>
<td>New York Zoo</td>
<td>This study</td>
<td>GR</td>
</tr>
<tr>
<td><em>S. brevicollis</em></td>
<td>1903</td>
<td>New York Zoo</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>S. heterothallis</em></td>
<td>2738</td>
<td>Texas</td>
<td>This study</td>
<td>PA</td>
</tr>
<tr>
<td><em>S. heterothallis</em></td>
<td>2739</td>
<td>Texas</td>
<td>This study</td>
<td>OR</td>
</tr>
<tr>
<td><em>S. sclerogena</em></td>
<td>2741</td>
<td>Ceylon</td>
<td>This study</td>
<td>GR</td>
</tr>
<tr>
<td><em>S. sclerogena</em></td>
<td>2740</td>
<td>Ceylon</td>
<td>This study</td>
<td>PA</td>
</tr>
</tbody>
</table>
FGSC, Fungal Genetics Stock Center. N. crassa, Neurospora crassa; N. discreta, Neurospora discreta; N. sitophila, Neurospora sitophila; N. intermedia, Neurospora intermedia; N. tetrasperma, Neurospora tetrasperma; N. pannonica, Neurospora pannonica; N. dodgei, Neurospora dodgei; S. brevicollis, Sordaria brevicollis; S. heterothallis, Sordaria heterothallis; S. sclerogena, Sordaria sclerogena. PA, OR and GR refer to het-c$^{PA}$, het-c$^{OR}$ and het-c$^{GR}$ allelic specificity type as designated in Neurospora crassa (Saupe and Glass, 1997). X, unidentified specificity.
2.2.3 Polymerase Chain Reaction (PCR) amplification of het-c specificity region

Genomic DNA was used for PCR amplification. The region of het-c allelic specificity was amplified with primers designed from the sequence of the het-cOR allele (Saupe et al., 1996b). Positions of the primers in the het-c sequences are shown in Figure 2-3. The primers are: Red, 5'(598)GGAGACATGGCGATATCG(615)3'; Yellow, 5'(1441)GTGAGGCACAACCCACTC(1424)3'. PCR reactions were prepared in 25 μl volumes including 100 ng of genomic DNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM each of dATP, dTTP, dCTP and dGTP, 5 pmol of each primer, and 2.5 unit of Taq DNA polymerase (Boehringer Mannheim, Indianapolis, IN). Amplification cycles were performed under the following conditions: 95°C for 5 min followed by 30 step cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, and 72°C for 10 min after the last cycle.

2.2.4 Subcloning of PCR products

PCR products were subcloned into the pCRII vector using the TA cloning kit (InVitrogen, San Diego, CA). Ligation reactions were performed according to the manufacturer’s guide. Subcloning efficiency competent Escherichia coli DH5α cell (F, endA1, hsdR17, supE44, lacZm15) was used for transformation. Plasmids were prepared using Qiagen miniprep kits (Qiagen, Chatsworth, California).
Figure 2-3. Diagram of the primer locations for PCR amplification of het-c specificity domain (SP). The amino acid sequences of the polymorphic region are shown as well. OR, HET-c^{OR}. PA, HET-c^{PA}. GR, HET-c^{GR}. 

EALRCLGQALHTLEDFPAPHSNYCELVLIDMEERRGHH--SPVFPHVGTDRTRITLRNDTRNNNG--KSVWPLVTGTFGG OR
EALRCLGQALHTLEDFPAPHSNYCELALIDIEKETRSESRLIFPHVGTATRITL--NNG--KLVWPLVTGTFGG GR
EALRCLGQALHTLEDFPAPHSNYCELVLIDMEERRGHH--SPVFPHVGTATKLKL--ENRQFRRVRFGEYDSGAYAQPPLVTGTFGG PA
The DNA sequence of PCR products was determined using the ABI Taq DyeDeoxi Terminator cycle method (Mississauga, ON) on an ABI 373 automatic sequencer at the NAPS (Nucleic Acid and Protein Service) unit, Biotechnology Laboratory, University of British Columbia. Approximately 500 ng of purified plasmid containing the expected PCR product was used as a template in a total of 20 μl reaction mixture containing 3.2 pmol primer (Red or Yellow; see section 2.2.3). The sequencing reaction was performed in a Perkin-Elmer thermal cycler for a total of 25 cycles as follows: 96°C for 30 sec, 50°C for 15 Sec, 60°C for 4 min.

2.2.6 Sequence analysis

The DNA sequences were translated into amino acid sequences using universal genetic codons with the program MacVector 4.1 (International Biotechnologies, Inc., Tecnomara, Fernwald, Germany), aligned using ClustalW 6.1 (Institute for Biomedical Computing, Washington University, St. Louis, MO), and adjusted manually according to positional homology. DNA sequences of the het-c specificity region from the isolates listed in Table 2-1 were also aligned by ClustalW 6.1 and modified manually according to the corresponding amino acid sequence positional homology. Insertions or deletions were positioned according to the amino acid alignments.

2.2.7 Phylogenetic tree reconstruction

The phylogenetic tree of het-c was constructed using PHYLIP 3.0 (Felsentein, 1993). Three different methods, the maximum parsimony method (Swofford, 1990), the neighbor-
joining method (Saitou and Nei, 1987) and the maximum likelihood method (Felsenstein, 1981), were used for tree reconstruction.

Parsimony methods search for the tree with the fewest evolutionary changes (e.g., the smallest number of nucleotide changes in the case of DNA sequence data) or steps (transformations from one character state to another) and are not based on explicit evolutionary models (i.e., there are not specific parameters or assumptions for nucleotide substitution probabilities) for computation. Since no explicit evolutionary model is assumed with parsimony, the phylogenetic trees can be inconsistent (Felsenstein, 1988). However, the computation time is short with parsimony methods.

The neighbor-joining method is based on distance calculation between species (Saitou and Nei, 1987). For nucleotide sequence data, the distances are calculated from the fraction of sites differing between two sequences and using specified nucleotide “transversion vs. transition ratio” corrections. The algorithm makes a prediction of the distance for each pair as the sum of branch lengths in the path from one species to another through the tree. Neighbor-joining does not check the possibility of alternative trees, thus it is difficult to see how different kinds of discrepancies from a tree are weighted. The algorithm is very fast, it is particularly effective for large studies or for bootstrap resampling studies.

Maximum likelihood methods produce consistent results if the model is correct. However, with many taxa, it is difficult to search exhaustively for the maximum likelihood tree. Maximum likelihood can estimate the parameters (e.g., transversion/transition rate, variation of substitution rate from site to site) of the model from the data. Maximum likelihood provides a way of handling missing or ambiguous data, as the probability of observing each of the bases allowed by the ambiguity can be explicitly evaluated. Maximum
likelihood considers that nucleotide changes are more likely along long branches than short ones and estimation of branch lengths is an important component. The method of computing maximum likelihood can be very slow, and large amounts of computation time are required for trees with many species or bootstrapping.

In parsimony tree reconstruction, gap sites were treated as a fifth character, with a deletion or insertion motif treated as a single event. In neighbor-joining tree reconstruction, distances were estimated using the Dnadist program in PHYLIP 3.0 with Kimura's two-parameter correction (Kimura, 1980), and gap sites were eliminated in all pairwise distance calculations. In maximum likelihood tree reconstruction, regions of deletion were treated as “missing data”. The robustness of the dendrogram topology was determined by bootstrap with 500 replicates for both parsimony and neighbor-joining trees (Felsenstein, 1985).

2.2.8 Calculation of synonymous and nonsynonymous nucleotide substitutions

Number of nonsynonymous substitutions per nonsynonymous site ($d_N$) and synonymous substitutions per synonymous site ($d_S$) was calculated by pairwise comparison of sequences from all isolates with the program MEGA (Kumar et al., 1993) according to Juke-Cantor's correction. Gap sites were treated according to the “pairwise-deletion” option in the MEGA program, which ignores the gaps present in pairwise comparisons of sequences. Microsoft Excel was used to map the substitutions to each codon position.
2.2.9 Statistical analyses

A standard z-score test (Capon, 1988) was used to test the following hypotheses at the P<0.05 significant level: 1) mean $d_s = \text{mean } d_N$; 2) mean $d_N$ in $het-c$ polymorphic region (the region of the 34 – 48 amino acids which confers $het-c$ allelic specificity) = mean $d_N$ in $het-c$ non-polymorphic region; 3) the ratio of $d_N$: $d_s$ in the $het-c$ polymorphic region = the ratio of $d_N$: $d_s$ in the $het-c$ non-polymorphic region. Values of $d_N$: $d_s$ ratios and the associated standard errors were calculated using the delta method (Kotz and Johnson, 1989; Haydon et al., 1998).
2.3 RESULTS

2.3.1 Genomic Southern blot analysis

To examine if the DNA sequence of het-c is conserved in the Sodariaceae, genomic DNA from 15 isolates, including N. crassa, N. discreta, N. sitophila, N. intermedia, S. sclerogenia and S. brevicollis, was subjected to Southern blot analysis probed with a 3.9 kb fragment of the het-c$^{OR}$ genomic clone. Under conditions of both low and high stringency, only a single band corresponding to het-c was detected in the genome of all the isolates analysed (Figure 2-4). This result indicates that het-c is highly conserved in different Neurospora and even Sordaria species, and that het-c is present in a single copy in these isolates, as was previously determined for N. crassa strains (Saupe et al., 1996).

2.3.2 PCR products and DNA sequences

To examine evolutionary features of het-c, genomic DNA from all the isolates was amplified by PCR using primers spanning the het-c specificity region of N. crassa (Figure 2-3) and the PCR products were sequenced. A single band was produced in each PCR reaction with the expected size of 830 bp to 860 bp (data not shown). The DNA sequences of the het-c specificity region differed among these isolates (see Appendix 7.1), even within a single species. The two most distantly related alleles of all the isolates, Np7221 and Sh2739, showed only 66% nucleotide identity.
Figure 2-4. Southern hybridization of genomic DNA from different species and genera (listed in Table 2.1) to a plasmid carrying the \( \text{het-c}^{OR} \) clone from \( \text{N. crassa} \). Genomic DNA was digested with \( \text{Sac I} \) and \( \text{Pst I} \). Sequences corresponding to \( \text{het-c} \) are apparently present as a single copy in all tested isolates.
2.3.3 Amino acid sequence of the het-c specificity region

Although the DNA sequences of the het-c specificity region varied among the isolates, the inferred amino acid sequences from these isolates fell into one of the three HET-c specificity types based on the pattern of insertion/deletions as first observed in the N. crassa het-c alleles (Figure 2-5, Saupe and Glass, 1997). A comparison of the HET-C peptide sequences from all the isolates showed conserved and variable domains. Regions flanking the variable domain showed a high level of amino acid identity among all the isolates. Two amino acid blocks, I and II, are highly polymorphic among het-c alleles (Figure 2-5). The polymorphic block I contains the sequences of MEERRGG(Q)H and IH(Q)EKEET(D)R(G/C)S that are found in both HET-c^PA and HET-c^OR groups. For all species containing a HET-c^GR - type allele, the polymorphic block I contains only a similar sequence of IHEK(N)ET(N)R(C,P)S. Partial duplication analysis of the two N. crassa strains, Nc1824 and Nc2489, indicated that amino acid sequence composition in the polymorphic block I does not affect het-c specificity (Saupe and Glass, 1997). The second polymorphic block (II) contains the sequences that are divergent among all het-c allelic types. HET-c^OR - type alleles contain a sequence motif of TA(V,D)TRV(I)TL in block II, while HET-c^GR - type alleles contain a sequence motif of TA(V)TR(Q)ITL, with the exception of Ndo1692, which contains the sequence GAHTRKTL. The polymorphic block II in HET-c^PA - type of alleles contain sequences of TA(E,R)TKKLK(V,R) or AHTLR(K,T)Y(L). Chimeric studies (See chapter 3) demonstrated that the sequence diversity in the polymorphic block II also does not affect het-c specificity. A five amino acid block, FPHVG, located between the two polymorphic blocks is conserved among all the isolates.
Across species and alleles, HET-c peptide sequence variation in the specificity motif (Figure 2-5) occurred at 43 out of 91 amino acid residues. Within HET-c^{OR} - type alleles, the variation occurred at fourteen out of 74 amino acid residues. Within HET-c^{GR} - type alleles, the variation occurred at thirteen out of 70 amino acid residues. Within HET-c^{PA} - type alleles, the variation occurred at 41 out of 85 amino acid residues. The fifteen to sixteen amino acid sequences within the het-c^{PA} specific insertion region IR2 (Figure 2-5) are the most polymorphic. An exceptional allele, Ndi5923, was found in N. discreta, which has a deletion of five amino acid residues in the het-c^{PA} - type specific insertion motif.

With the exception of the pseudohomothallic species, N. tetrasperma, at least two of the three het-c allelic variants were found in all species with more than one isolate. As shown in Figure 2-5, in the heterothallic species, S. brevicollis, the two inter-fertile isolates from the New York Zoo each contained a het-c allele of a different HET-c type, HET-c^{OR} (Sb1903) and HET-c^{GR} (Sb7140). The two S. sclerogena inter-fertile isolates from Ceylon also have different HET-c allelic types, one showing HET-c^{PA} type (Ss2740) and the other showing HET-c^{GR} - type (Ss2741). Two different HET-c allelic types, HET-c^{OR} (Sh2739) and HET-c^{PA} (Sh2738) were also found in the two inter-fertile S. heterothallis isolates. Similarly, two alternative het-c allelic types, HET-c^{PA} and HET-c^{OR}, were found in the four inter-fertile Gelatinospora sp. isolates (Gsp 8243, Gsp8241, Gsp8239 and Gsp8242). At least two or all three of the het-c allelic types were identified in the heterothallic Neurospora species, N. crassa (Nc1120, Nc4711, Nc4732, Nc4486, Nc1945, Nc1455, Nc967, Nc4499, Nc4481, Nc1824, Nc2489, Nc847, Nc4709, Nc1693), N. sitophila (Ns5941, Ns 963, Ns5940), N. intermedia (Ni2316, Ni3290, Ni3721, Ni1940, Ni1799), and N. discreta (Ndi3228, Ndi5923, Ndi3268, Ndi6788, Ndi6793).
In summary, an amino acid sequence comparison of the *het-c* specificity region from all of the 40 isolates showed that two to five different amino acid residues were found at each polymorphic site from different isolates in the HET-c specificity domain, in particular within the two polymorphic blocks and the HET-c$^{PA}$ - specific insertion motif. The data suggests that polymorphism in the *het-c* specificity region is shared among different *Neurospora* species and the three genera (*Neurospora, Sordaria* and *Gelasinospora*) in the Sordariaceae.
Figure 2-5. Inferred amino acid sequences of PCR amplified het-c allelic specificity motif from 40 isolates. The alignment was obtained by ClustalW 6.1 and modified manually. Ne, Neurospora crassa; Ndi, Neurospora discrete; Ns, Neurospora sitophilia; Nt, Neurospora tetrasperma; Np, Neurospora pannonica; Ndo, Neurospora dodgei; Ss, Sordaria sclerogenia; Sh, Sordaria heterothallis; Sb, Sordaria brevicollis; Gsp, Gelasinospora sp. Numbers following indicate the Fungal Genetics Stock Center (FGSC) number or P number (4481, 4486, 4499 strains). PA (Panama), GR (Groveland), and OR (Oakridge) refer to the three molecularly characterized alleles that confer different het-c specificity (Saupe and Glass, 1997). Asterisks and dashes indicate identical and deletion residue sites, respectively. Underlined section is het-c specificity variable region. Polymorphic block I and II, putative OR-specific insertion sequence (IS1) and PA-specific insertion motif (IS2) are indicated.
2.3.4 Trans-specific polymorphism of het-c

Sequence analysis of the PCR products of the het-c specificity region from 40 natural isolates showed 39 distinct alleles based on nucleotide sequences (Nc1130 and Nc4711 are identical; Appendix 7.1); 24 alleles had distinct amino acid sequences (Figure 2-5). In order to examine the relationship of the het-c alleles, I performed a phylogenetic analysis of DNA sequences of the het-c specificity region using the neighbor-joining algorithm based on Kimura's two-parameter distances (Figure 2-6, Figure 2-7), parsimony (Figure 2-8, Figure 2-9), and maximum-likelihood (Figure 2-10, Figure 2-11). Phylogenetic trees were reconstructed under two conditions: including and excluding the allelic specific insertion/deletion sequences (sequences coding for IS1 and IS2 in Figure 2-5). Thirty-one sequences isolated from seven Neurospora species (N. crassa, N. discreta, N. intermedia, N. sitophila, N. tetrasperma, N. pannonica and N. dodgei), five sequences from three Sordaria species (S. brevicollis, S. heterothallis and S. sclerogenia) and three sequences from Gelasinospora sp. were analysed.

Phylogenetic analysis showed that het-c subgrouped by allelic type rather than according to species - a trans-specific mode of polymorphism. In all the trees reconstructed using DNA sequences that include the insertion/deletion sites in the het-c specificity region, the clustering pattern of het-c alleles was supported by robust bootstrap values. Five hundred bootstrap replicates were analysed using both neighbor-joining (Figure 2-6) and parsimony (Figure 2-8) methods. Eighty-five percent of neighbor-joining trees and 100% of parsimony trees supported the clustering of the eleven het-c\textsuperscript{OR} - type alleles, Nc2489, Nc1693, Nc4709, Nc847, Nc4481, Gsp8242, Sh2739, Ni1940, Ni1799, Ns963 and Ns5940. All the trees (100%) constructed using neighbor-joining and parsimony methods supported the clustering of the
four het-c\textsuperscript{OR} - type alleles, \textit{Sb1903}, \textit{Nd16793}, \textit{Gsp8239} and \textit{Gsp8241}. Ninety - three percent of neighbor-joining trees and 99% of parsimony trees supported the subgrouping of the nine het-c\textsuperscript{PA} - type alleles, \textit{Ndi3228}, \textit{Nc4486}, \textit{Nc4711}, \textit{Nc1130}, \textit{Nt1270}, \textit{Nt6583}, \textit{Ni2316}, \textit{Ni3290} and \textit{Nc4832}. Ninety-five percent of neighbor-joining trees and 92% of parsimony trees supported the clustering of the three het-c\textsuperscript{PA} - type alleles, \textit{Ss2740}, \textit{Np7221} and \textit{Gsp8243}. Eighty-percent of the neighbor-joining trees and 87% of the parsimony trees supported the subgrouping of seven het-c\textsuperscript{GR} - type alleles, \textit{Nc967}, \textit{Nc4499}, \textit{Nc1455}, \textit{Ss2740}, \textit{Sb7140}, \textit{Nd16788} and \textit{Nc1945}. A similar grouping of het-c alleles was also observed in the tree reconstructed using maximum likelihood (Figure 2-10). In each of the groups, alleles were found from at least two species.

It is possible that the clustering of het-c alleles is affected by the DNA sequence alignment bias based on the placement of the insertion/deletion sites. To examine this possibility, trees were reconstructed using DNA sequences excluding the deletion/insertion sites. Although certain differences in the subgroup branching points and weaker bootstrap support were observed, the subgroupings of het-c alleles were consistent in all trees. In 500 neighbor-joining trees (Figure 2-7), 99% of the replicates supported the subgrouping of the fifteen het-c\textsuperscript{OR} - type alleles from seven species and three genera and the subgrouping of the ten het-c\textsuperscript{PA} - type alleles from five \textit{Neurospora} species; 92% of the replicates supported the subgrouping of the five het-c\textsuperscript{GR} - type alleles from three species; and 94% of the replicates supported the subgrouping of the three het-c\textsuperscript{PA} - type alleles, \textit{Gsp8243}, \textit{Ss2740} and \textit{Np7221}. In 500 parsimony trees (Figure 2-9), 85% of the replicates supported the subgrouping of the three het-c\textsuperscript{PA} - type alleles, \textit{Gsp8243}, \textit{Ss2740} and \textit{Np7221}; 98% of the replicates supported the subgrouping of the four het-c\textsuperscript{OR} - type alleles \textit{Nd6793}, \textit{Sb1903}, \textit{Gsp8241} and \textit{Gsp8239}.
However, weak bootstrap supports (below 70%) were observed in the following clusterings: the subgrouping of five *het-c* GR-type alleles, the lineage separation of the ten *het-c* PA-type alleles and the eleven *het-c* OR-type alleles (Figure 2-9). Nonetheless, the subgrouping of the nine *het-c* PA-type alleles from four species was supported by 87% of the bootstrap replicates (Figure 2-9). Most of the *het-c* allele groups were also observed in the tree reconstructed using maximum likelihood (Figure 2-11).

The data indicate that the subgroupings of *het-c* by allelic types were not significantly affected by the DNA sequence alignments. These analyses suggest that the DNA sequences of *het-c* specificity region display trans-specific polymorphism; alleles from the same allelic types are more similar than alleles from different allelic types but from the same species. For example, in all the trees, a *het-c* PA-type allele from *N. crassa* (Nc4832) is more closely related to another *het-c* PA-type allele from *N. intermedia* (Ni3290) than to a *het-c* OR-type allele from *N. crassa* (Nc847).

An interesting finding is that the trans-specific polymorphism of *het-c* is also apparent within each allelic group. For instance, in 500 neighbor-joining trees constructed using DNA sequences from only *het-c* PA-type alleles (Figure 2-12), 97% of the bootstrap replicates supported the trans-specific clustering of Nc4832, Ni2316 and Ni3290, and the subgrouping of Nt6583, Nc4711, Nc1130, Nt 1270, Ndi3228 and Nc4486; all of the replicates supported the clustering of Gsp 8243, Np7221 and Ss2740. Nearly identical clustering patterns were observed in trees constructed using maximum likelihood method (data not shown). Trans-specific lineages are also observed within both *het-c* OR-type alleles and *het-c* GR-type alleles. In 500 neighbor-joining trees of *het-c* OR-type alleles (Figure 2-13), 82% of the bootstrap replicates supported the clustering of *N. crassa* allele Nc1824 with *N. intermedia*
allele Ni3721, and the subgrouping of Ndi6793, Sb1903, Gsp8241 and Gsp8239; 84% of the bootstrap replicates supported the clustering of alleles Nc4481 and Gsp8242. In 500 neighbor-joining trees of het-c\textsuperscript{GR} - type alleles (Figure 2-14), all the bootstrap replicates (100%) supported the subgrouping of N. crassa allele Nci945 with N. discreta allele Ndi6788 and the subgrouping of Nc967, Ss2741, Sb7140 and Ndo1692.
Figure 2-6. Unrooted neighbor-joining tree from the DNA sequence (390 nucleotides) alignment in the *het-c* specificity region including the insertion/deletion sites. Bootstrap support is shown by percentage out of 500 replicates. Percentages below 75% are not indicated. Branch lengths are proportional to the genetic distances. The designation OR, *het-c*\textsuperscript{OR} allelic-type; PA-*het-c*\textsuperscript{PA} allelic type; and GR-*het-c*\textsuperscript{GR} allelic type is based on the alignment in Figure 2-5 and is used in all the following Figures.
Figure 2-7. Unrooted neighbor-joining tree from the DNA sequence (327 nucleotides) alignment in the het-c specificity region excluding the insertion/deletion sites. Bootstrap support is shown by percentage out of 500 replicates. Branch lengths are proportional to the genetic distances.
Figure 2-8. Unrooted parsimony tree from the DNA sequence (390 nucleotides) alignment in the *het-c* specificity region including the insertion/deletion sites. Bootstrap support is shown by percentage out of 500 replicates.
Figure 2-9. Unrooted parsimony tree from the DNA sequence (327 nucleotides) alignment in the het-c specificity region excluding the insertion/deletion sites. Bootstrap support is shown by percentage out of 500 replicates.
Figure 2-10. Unrooted maximum likelihood tree from the DNA sequence (390 nucleotides) alignment in the *het-c* specificity region including the insertion/deletion sites.
Figure 2-11. Unrooted maximum likelihood tree from the DNA sequence (327 nucleotides) alignment in the \textit{het-c} specificity region excluding the insertion/deletion sites.
Figure 2-12. Unrooted neighbor-joining tree of $het-c^{PA}$-type alleles. Bootstrap support for a clade is shown as the percentage of 500 replicates containing the clade. Percentages below 75% are not indicated. The sequence length is 375 nucleotides.
Figure 2-13. Unrooted neighbor-joining tree of \( \text{het-c}^{\text{OR}} \) - type alleles. Bootstrap support for a clade is shown as a percentage of 500 replicates. Percentages below 75% are not shown. Note that neither species nor genera are monophyletic. The sequence length is 345 nucleotides.
Figure 2-14. Unrooted neighbor-joining tree of \textit{het-c}^{GR} - type alleles. An \textit{N. crassa} allele \textit{Nc967} is more closely related to a \textit{S. sclerogenia} allele \textit{Ss2741} than to \textit{N. crassa} alleles \textit{Nc4499} or \textit{Nc1455}. Bootstrap is shown as a percentage of 500 replicates (only > 60\% are shown). Again, alleles from the same species or genera are not monophyletic. The sequence used is 327 nucleotides.
2.3.5 Numbers of synonymous and nonsynonymous substitutions per site in the *het-c* specificity region.

Theoretical studies predict that positive Darwinian selection should enhance the number of nonsynonymous nucleotide substitutions (Lopez de Castro et al., 1982). To test if positive selection was acting in favor of polymorphisms in the *het-c* specificity domain, the number of nonsynonymous substitutions per nonsynonymous site \((d^N)\) and synonymous substitutions per synonymous site \((d_S)\) in the polymorphic (filled bar region in Figure 2-15, as highlighted in Figure 2-5, total 57 codons) and non-polymorphic (clear region in Figure 2-15, total 73 codons) region of 130 codons was examined. Histograms of how the per-site rates of change mapped onto the codons for both synonymous and nonsynonymous substitutions are shown in Figure 2-15.

In all 40 *het-c* alleles, the ratio of \(d^N/d_S\) in the polymorphic region (20.4 / 38 = 0.54) significantly exceeded that in the non-polymorphic region (2.6 / 24.5 = 0.11; \(z\) test, \(p < 0.005\); Table 2-3) due to the significantly increased value of \(d^N\) in the polymorphic region. A significantly increased \(d^N/d_S\) ratio in the polymorphic region was also observed in the separate analyses of *het-c* alleles from *het-c*\(^{OR}\) - type, *het-c*\(^{PA}\) - type and *het-c*\(^{GR}\) - type, *N. crassa* isolates and the genus *Sordaria* (Table 2-2). The most variable alleles contained the *het-c*\(^{PA}\) - type insertion, where the \(d^N/d_S\) ratio reached 1.3 in the polymorphic region.

In the estimation of \(d^N\) and \(d_S\) among all *het-c* alleles, the \(d_S\) value was generally higher than the value of \(d^N\) with the exception of the polymorphic region of PA-type alleles (Table 2-2). However, the value of \(d^N\) was significantly higher in the polymorphic region than that in the non-polymorphic region in all analyses (Table 2-2, \(z\) test, \(p < 0.005\)) while the \(d_S\) values in the two regions showed no significant differences (\(z\) test, \(p < 0.05\)). Thus, the
ratio of $d_\nu/d_s$ in the polymorphic region was significantly greater than that of the non-polymorphic region (Table 2-2).
Figure 2-15. a) The average number of nonsynonymous nucleotide substitutions per nonsynonymous site and b) the average number of synonymous nucleotide substitutions per synonymous site grouped by codon of het-c specificity domain and its flanking region. The filled bar region represents the polymorphic region of het-c as highlighted in Figure 2-5. Clear bar region represents the flanking region. Codon position 40 corresponds to the position of the first amino acid residue in Figure 2-5. The 30 amino acids from codon 40 to 60 and from codon 120 to 130 are identical in all alleles. The numbers of nonsynonymous and synonymous substitutions were calculated by MEGA (Kumar et al., 1993). All 40 alleles were used in the analysis. Note the difference in ordinate scales used for d_N and d_S.
Table 2-2. Comparison of number of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per 100 sites in *het-c* polymorphic (PM) and non-polymorphic (Non-PM) region.

<table>
<thead>
<tr>
<th>Alleles/Species</th>
<th>PM Region</th>
<th>Non-PM Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d_S</td>
<td>d_N</td>
</tr>
<tr>
<td>All***</td>
<td>38.0</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td>(7.1)</td>
<td>(2.8)</td>
</tr>
<tr>
<td>OR-type*</td>
<td>25.5</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>(5.6)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>GR-type*</td>
<td>16.2</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>(5.1)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>PA-type***</td>
<td>16.3</td>
<td>21.5</td>
</tr>
<tr>
<td></td>
<td>(4.5)</td>
<td>(2.9)</td>
</tr>
<tr>
<td><em>N. crassa</em>*</td>
<td>23.9</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>(7.8)</td>
<td>(3.3)</td>
</tr>
<tr>
<td><em>Sordaria</em>**</td>
<td>37.0</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>(8.9)</td>
<td>(3.8)</td>
</tr>
</tbody>
</table>

Polymorphic (PM) and non-polymorphic (non-PM) regions refer to the codons of filled (PM) and clear (non-PM) bar regions in Figure 2-5, respectively. N gives the number of alleles surveyed. Standard errors are in brackets. OR, GR and PA-type refer to *het-c*\(^{OR}\), *het-c*\(^{GR}\) and *het-c*\(^{PA}\) allelic specificity types, respectively. Test of hypothesis: d_N/d_S ratio has no significant difference in the two regions. * P < 0.05, ** P < 0.01, *** P < 0.001.
2.4 DISCUSSION

2.4.1 Trans-specific polymorphisms at *het-c*

Three methods were used for *het-c* phylogenetic inference from DNA sequence data. Trans-specific polymorphisms were observed for *het-c* alleles using all three methods. Regardless of whether the insertion/deletion motifs were included or excluded, alleles clustered according to allelic types rather than according to species. In cases of including and excluding the insertion/deletion sites in the DNA sequences, maximum likelihood analysis resolved nearly identical subgroups according to *het-c* allelic types (Figure 2-10 and Figure 2-11); the neighbor-joining also gave similar subgroupings with similar robustness of bootstrap support (Figure 2-6 and Figure 2-7). Most of the subgroups in parsimony trees were similar with strong bootstrap support regardless of whether the insertion/deletion motifs were included or excluded (Figure 2-8 and Figure 2-9).

The *het-c* region that confers allelic specificity displayed a high degree of polymorphism in DNA sequences. Thirty-nine *het-c* alleles were identified in the 40 Sordariaceae sequences examined and fifteen were identified in *N. crassa* alone. In the phylogenetic analyses of the 39 alleles, *het-c* formed lineages according to allelic types rather than species, suggesting that balancing selection is operating to maintain the *het-c* polymorphisms. Interestingly, the trans-specific polymorphisms are not limited to the three *het-c* allelic types, defined as *het-c*<sup>QR</sup>- type, *het-c*<sup>GR</sup>- type and *het-c*<sup>PA</sup>- type, but are also observed within an allelic type (Figure 2-12 - 2-14).

Trans-specific polymorphisms have been well demonstrated at the *MHC* and the *S loci*, both of which are believed to be under balancing selection and function as a self and
non-self recognition system. At the MHC loci, certain HLA alleles are more similar in their sequence to certain chimpanzee MHC alleles than to other HLA alleles (Lawlor et al., 1988; Mayer et al., 1988). At the S-loci, the trans-specific pattern of polymorphism has been demonstrated in both gametophytic and sporophytic self-incompatible systems (Dwyer et al., 1991; Ioerger et al., 1990; Richman et al., 1996).

Trans-specific polymorphism is not restricted to loci under balancing selection, such as the MHC loci and the S loci, but by the persistence of neutral polymorphisms if speciation times are short (in terms of generations) and if the effective population size is too large (Nagal et al., 1998). A well studied example of neutral trans-specific polymorphisms is represented by genes in the haplochromine species flock in Lake Victoria that underwent rapid diversification. In Lake Victoria cichlid fish, neutral trans-specific polymorphisms have been detected in the non-coding sequences of 3'UTR or introns of randomly selected nuclear genes and also in association with structural genes under purifying selection (e.g., G6P, a gene encoding glucose-6-phosphatase) (Ngal et al., 1998; reviewed in Klein et al., 1998). Most of the polymorphisms were estimated to have arisen more than 12,000 and less than two million years ago. It was proposed that the neutral polymorphisms in Lake Victoria species were passed on to the descendent species as they emerged during the explosive diversification and speciation events that created the extant flock (reviewed by Klein, 1998). However, the Victoria cichlid fish represents an extreme case of rapid speciation, where the maintenance of polymorphisms across speciation events may be expected. The divergence among the Sordariaceae examined here are thought to be much older (Berbee and Taylor, 1992; Berbee et al., 1995).
2.4.2 *het-c* allelic lineages were generated prior to the divergence of modern species

Sequence comparison and phylogenetic analysis of the *het-c* specificity region revealed the following associations: 1) the *het-c* sequences are highly divergent between allelic types and between alleles within each allelic type (Appendix 7.1); 2) *het-c* shows ubiquitous trans-specific polymorphism between allelic types and within allelic types (Figure 2-6 to Figure 2-14); 3) allele lineages were generated before the divergence of *Neurospora* and *Sordaria* and persisted through multiple speciation events.

The divergence time of *Neurospora* and *Sordaria* was calculated according to the 18S rRNA sequences, for which the molecular clock has been calibrated using fossil evidence and ages of fungal host and symbionts (Berbee et al., 1995). The DNA sequences of the 18S rRNA genes from *Sordaria fimicola* (Genebank access No. X69851) and *N. crassa* (Berbee and Taylor, 1992) show a 1.46% difference, indicating a 0.73% substitution per lineage. The calibrated nucleotide substitution rate is 0.5 - 2% per 100 million years in fungi (Berbee et al., 1995). Thus, *Sordaria* diverged from *Neurospora* at least 36 million years ago, indicating that the *het-c* polymorphism has been retained for at least this time interval. Long persistence of allelic lineages was also observed in both the MHC and S loci. At the DRB1 locus, allelic lineages have been retained for at least 30 million years (Ayala et al., 1994), while the oldest allelic lineages diverged approximately 55 million years ago (Martin, 1993; Satta et al., 1996a, 1996b). At the S loci, allelic lineages have been retained for at least 36 million years, while the oldest allelic lineages diverged more than 70 million years ago (Ioerger et al., 1990). The persistence of polymorphism within the *het-c* specificity domain within the different genera during such a long period of time strongly implies that allelic variants have been maintained by balancing selection across multiple speciation events.
2.4.3 het-c allelic lineages

Although only three allelic specificities have been identified in *N. crassa* (*het-c*\(^{OR}\), *het-c*\(^{GR}\) and *het-c*\(^{PA}\)), at least seven distinctive allelic lineages were resolved in most of the *het-c* phylogenetic trees—three *het-c*\(^{OR}\) lineages, two *het-c*\(^{PA}\) lineages and two *het-c*\(^{GR}\) lineages. In all the neighbor-joining trees and maximum likelihood trees, two *het-c*\(^{OR}\) - type alleles, *Ni3721* and *Nc1824*, were included in the *het-c*\(^{GR}\) lineage. The clustering of these two OR-type alleles with GR-type alleles was probably due to the affect of the intragenic (interallelic) recombination event involving the polymorphic block MEERGGH/IHEKETRS (Figure 2-5). Similarly, *Gsp8243*, *Np7221*, and *Ss2740* were separated from the PA-lineage probably because of recombination between the polymorphic block I, MEERGGH / IH(Y)KK(N)ETGE(R). The inclusion of *Ndo1692* in the PA-lineage (*Gsp8243*, *Np7221*, and *Ss2740*) is also due to the effect of the polymorphic block I. The formation of a separate OR-lineage that is made of *Sb1903*, *Ndi6793*, *Gsp8239* and *Gsp8241* is due instead to sequence variation in polymorphic block I and II in these alleles (Figure 2-5, Appendix 7.1).

Three factors could explain the occurrence of seven *het-c* allelic lineages. First, gradual accumulation of spontaneous nucleotide substitutions resulted in allelic lineages that arose at different evolutionary time intervals. Second, segmental sequence exchanges could have occurred between two lineages through interallelic (lineage) recombination. Consistent with this explanation, apparent mosaic sequences between *het-c*\(^{GR}\) - type and *het-c*\(^{OR}\) - type, plus *het-c*\(^{GR}\) - type and *het-c*\(^{PA}\) - type alleles were observed. For instance, *Ni3721* and *Nc1824* are chimeras between HET-c\(^{GR}\) (the polymorphic block IH(Y)EKETG(R)S(E)) and HET-c\(^{OR}\) (the OR-specific pattern of insertion RNDTR); *Np7221*, *Gsp8243*, *Ss2740* and *Ndo1692* are chimeras between HET-c\(^{GR}\) (the polymorphic block...
IHEKETGS/IYE(K)NEN(T)PRR) and HET-c PA (PA-specific pattern of insertion). Under this hypothesis, the polymorphic block in het-c would have a strong effect on the placement of these chimeric alleles in phylogenetic tree reconstructions, and thus additional het-c allelic lineages were formed. However, all lineages showed a trans-specific pattern, suggesting that generation of these exchanges was in a common ancestor of these or that recombination has happened more than once. Intragenic (interallelic) segmental exchanges were also observed in Feline MHC class I DRB genes based on the nature of a highly mosaic structure among the DNA sequences (Yuhki and O'Brien, 1990; 1994). Third, selection was acting on the functional region of HET-c in relationship to its role for non-self recognition. Negative (purifying) selection acted to conserve certain residues in regions required for common HET-c function; positive selection acted to produce sequence variations in regions that are important for HET-c specificity. If this is true, do the het-c allelic lineages in the phylogenetic analyses represent different functional het-c lineages that confer alternative specificities? It is known that the three het-c allelic types, het-c OR - type, het-c PA - type and het-c GR - type are mutually incompatible in natural heterokaryons (Garnjobst, 1953; Mylyk, 1975; Perkins, 1988), laboratory genetic analyses (Saupe and Glass, 1997) and transformation assays in N. crassa (Chapter 3). However, only four het-c specificities were detected among the 39 alleles (allele Ndi5923 confers different specificity from het-c OR, het-c PA and het-c GR) by chimeric construction and transformation assay (Chapter 3). The fact that different phylogenetic lineages were observed within a het-c allelic - type permits the following speculation: strains carrying the same allelic - type of het-c but assigned in different lineages may be incompatible in nature, but laboratory assays are not discriminating enough to differentiate them.
2.4.4 Possible mechanisms of balancing selection at the *het-c* locus

Balanced allelic polymorphisms can be maintained either by over-dominant selection (heterozygotes have a superior fitness) in a diploid organism or by frequency-dependent selection (rare alleles have a selective advantage) (Takahata and Nei, 1990; Takahata et al., 1992). Both selection mechanisms have been proposed to occur in the *MHC* and the *S* loci for the maintenance of polymorphisms (for reviews, see Hughes and Yeager, 1998; Charlesworth and Awadalla, 1998; and also see Introduction). In haploid filamentous fungi, heterozygote superiority (overdominance) selection is an unlikely force for the persistence of *het-c* polymorphisms. However, the possibility of overdominant selection at *het-c* can not be ruled out until *het-c* is shown to be inactive in the transient sexual dikaryon phase. In a population analysis of 36 *Neurospora crassa* isolates from a 5-hectare sugar cane field of Lousinana, the three *het-c* allelic - types showed nearly equal frequency of occurrence (this study and Louise Wheeler, personal communication). In a restriction fragment length polymorphism (RFLP) analysis of the PCR products of the *het-c* specificity region from the 36 isolates, fourteen showed a *het-c*\(^{PA}\) pattern, fourteen had a *het-c*\(^{OR}\) pattern and eight had a *het-c*\(^{GR}\) pattern. When genomic DNA from the 36 isolates was digested and probed with a 35-kbp cosmid containing *het-c*\(^{OR}\), twenty-four different RFLP patterns were observed (Louise Wheeler, personal communication) indicating that the population was not clonal. These studies suggest that maintenance of *het-c* polymorphism within fungal populations is possibly attributable to frequency-dependent selection.
2.4.5 An excess of nonsynonymous nucleotide substitutions in the het-c polymorphic region

A powerful method of discriminating between selection and neutral polymorphism is to compare the number of synonymous and nonsynonymous nucleotide substitutions per site (Hughes and Nei, 1988, 1989). In the case of positive selection favoring diversity at the amino acid level, the nonsynonymous substitution rate is enhanced (Hughes and Nei, 1988). Under purifying selection, as occurs in the case of most structural protein-coding genes or functional constraint region of some other molecules, the rate of synonymous substitution is higher (Kimura, 1983).

Hughes and Nei (1988) hypothesized that natural selection is responsible for the increased nonsynonymous nucleotide substitution rates. Significantly enhanced nonsynonymous nucleotide substitution has been shown in the hypervariable regions of both the MHC molecule and S alleles. It was found that the nonsynonymous nucleotide substitution was significantly enhanced in the 57 codons encoding the PBR of the MHC class I molecules (Nei and Gojobori, 1986; Cadavid et al., 1997). A similar pattern was also found in the putative PBR in the class II MHC genes (Hughes and Nei, 1989; Yuhki and O'Brien, 1997; Miller and Withler, 1996). At the S loci, higher numbers of nonsynonymous nucleotide substitutions were observed in the receptor domains of SRK (S-Receptor Kinase) and SLG (S-Locus Glycoprotein) genes in the Brassiceae (Hinata et al., 1995) and the hypervariable regions of S alleles in the Solanaceae (Clark and Kao, 1991; Ioerger et al., 1990). The hypervariable regions of these molecules play a direct role in the recognition of self and non-self and are believed to be under balancing selection to maintain allelic polymorphisms.
To assess the effect that selection has on the pattern of nucleotide substitution in het-c, the number and frequency of synonymous and nonsynonymous nucleotide substitutions inside and outside of the het-c variable region were compared. In the hypervariable region of het-c, both the numbers and frequency of nonsynonymous nucleotide substitutions were significantly higher. Since the number of synonymous nucleotide substitutions per site \((d_s)\) is expected to reflect the neutral mutation rate (the fraction of neutral mutations at synonymous sites being close to 100%, Kumar et al, 1993), the nearly uniform \(d_s\) values in HET-c indicate nearly-equal neutral mutation rates inside and outside the HET-c polymorphic regions. Therefore, the enhanced number of nonsynonymous nucleotide substitutions \((d_n)\) in the polymorphic region is unlikely due to a higher mutation rate at these sites.

Examination of the pattern of nucleotide substitutions in the het-c specificity region indicates that positive selection has acted to enhance nonsynonymous nucleotide substitutions favoring the diversity of het-c at the amino acid level. This evidence supports the hypothesis that HET-c has a selectively important function functions for non-self recognition, as with the MHC molecule and the S protein.

2.4.6 The het-c locus shares evolutionary features with MHC and S loci

In higher eukaryotes, the loci in the MHC and the S system are examples for which balancing selection has caused the retention of ancestral polymorphisms – trans-specific polymorphism. Both the MHC and the S loci have a large number of alleles present within a single species. In human MHC (the HLA complex) at the most polymorphic class I (HLA-B)
and II (HLA-DRB1) loci, 150 and 156 alleles have been identified, respectively (Bodmer et al., 1997). At the S locus, 30 to 50 S-alleles are commonly found per population in the Solanaceae (Lawrence et al., 1993) and more than 100 S-alleles have been estimated in Brassica campestris (Nou et al., 1993). Due to the lack of a large number of Neurospora isolates from a defined population, it has been difficult to estimate the number of existing het-c alleles in natural populations (Glass, personal communication). However, 39 het-c alleles have been identified in the Sordariaceae and at least seven allelic lineages were resolved in the phylogenetic analyses.

The data in this study demonstrate that the het-c region conferring specificity shares several common evolutionary features with the MHC and S loci. First, alleles display large divergences in DNA sequence. Second, trans-specific polymorphisms are apparent. Third, nonsynonymous nucleotide substitutions are enhanced within the specificity region. Fourth, allelic lineages persist for a long period of time. These data suggest that het-c is subject to balancing selection. This study is consistent with the hypothesis that het-c acts as a locus important for non-self recognition in the Sordariaceae and that vegetative incompatibility is a bona fide non-self recognition system in filamentous fungi that limits heterokaryon formation between unlike individuals. Moreover, comparable evidence for the persistence of ancient polymorphisms at the het-c locus, mammalian MHC loci and plant S' loci reinforces the view that balancing selection is an evolutionary mechanism common to different non-self recognition systems.
Chapter 3  Molecular basis of \textit{het-c} allele specificity

3.1 INTRODUCTION

As was described in the general introduction, the \textit{het-c} locus encodes three allelic specificities termed \textit{het-c}^{OR}, \textit{het-c}^{GR} and \textit{het-c}^{PA} based on genetic studies (Saupe and Glass, 1997). Polypeptides encoded by the three alleles show 86\% amino acid identity and differ by variable domains. By chimeric construction between the three \textit{het-c} alleles, a 34 – 48 amino acid variable region that is dissimilar in HET-c^{OR}, HET-c^{GR} and HET-c^{PA} was determined to confer allelic specificity (Saupe and Glass, 1997; Figure 1-7 and Figure 1-8).

The variable domain of HET-c^{OR}, HET-c^{GR} and HET-c^{PA} differs in both amino acid sequences and the pattern of insertion/deletion. An examination of this region in related species and genera to \textit{N. crassa} in Chapter 2 showed that all 40 alleles surveyed fell into one of the three \textit{het-c} allelic types based on the pattern of insertion/deletion within the \textit{het-c} specificity region (Figure 1-8; Wu et al., 1998). However, within each of HET-c^{OR}, HET-c^{GR} and HET-c^{PA} allelic types, amino acid sequences in the variable region was highly diversified among alleles.

In this chapter, I address the question of whether the amino acid sequence of a particular \textit{het-c} specificity domain or the pattern of insertion/deletion is the primary factor for determining specificity. To differentiate these two possibilities, I take advantage of the natural amino acid variation and deletion/insertion motifs among the 39 unique alleles isolated from different genera and species described in Chapter 2. First, I addressed whether the \textit{het-c} specificity of these alleles, as assayed in \textit{N. crassa}, was affected either by amino
acid sequence and/or insertion/deletion pattern in the specificity domain. By chimeric construction and transformation assays, I observed that the pattern of insertion/deletion in the specificity domain is the most critical determinant for *het-c* allele specificity. Second, I further investigated how the pattern of insertion/deletion affects *het-c* specificity by construction of a number of artificial *het-c* alleles using PCR mutagenesis. The combination of amino acid sequences and insertion/deletion motifs of these artificial alleles was not observed in the survey of natural alleles. By this method, I was able to generate a number of novel *het-c* specificities.

The mechanism of allele specificity has been examined in several fungal non-self recognition systems by the construction of chimeric (or hybrid) alleles. In *P. anserina*, a single amino acid difference in the proteins encoded by the vegetative incompatibility locus, the *het-s* locus, was sufficient to confer allele specificity (Deleu et al., 1993). Such a situation was also observed at the *het-c* locus in *P. anserina* (Saupe et al., 1995). In *Ustilago maydis*, a region composed of 30 to 48 amino acid residues was identified that regulates allele specificity at the *b* mating locus (Yee and Kronstad, 1993), and artificial hybrid *b* alleles with novel specificity were generated by chimeric construction within this border region (Yee and Kronstad, 1998). In *Coprinus cinereus*, specificity of the homeodomain mating protein HD1 and HD2 was determined by the N-terminal 160 to 170 amino acids (Banham et al., 1995). In the case of the homeodomain mating proteins in *Schizophyllum commune*, named the Y and Z proteins, a region of 55 amino acids in the Y protein and 41 amino acids in the Z protein was defined by chimeric allele construction that determined the specificity (Wu et al., 1996). Overall, these studies defined important elements involved in specificity determination and mediating non-self recognition.
3.2 MATERIALS AND METHODS

3.2.1 Strains and media

A list of N. crassa strains used for transformation and heterokaryon tests is given in Table 3-1 and Table 3-2, respectively, along with their relevant genotypes. All strains were maintained on Vogel's media (Vogel, 1964) with required supplements.

*Escherichia coli* strain DH5a (F-, endA1, hsdR17, supE44, lacZM15; Bethesda Research Laboratory) was used for routine DNA manipulation work and was grown in LB medium (Sambrook et al., 1989).

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Genotypes</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2-2-9</td>
<td>het-6^{PA} het-c^{OR} thr-2 A</td>
<td>M. L. Smith</td>
</tr>
<tr>
<td>C9-2</td>
<td>het-6^{OR} het-c^{PA} thr-2 a</td>
<td>M. L. Smith</td>
</tr>
<tr>
<td>FGSC 2193</td>
<td>NM149 T(II;V), het-6^{OR} het-c^{GR} A</td>
<td>D. D. Perkins</td>
</tr>
<tr>
<td>C15-1a</td>
<td>het-6^{OR} het-c^{OR}; pan-2; arg-5; inl a</td>
<td>R. Todd</td>
</tr>
<tr>
<td>Xa-3</td>
<td>het-6^{OR} het-c^{PA}; pan-2; arg-5 A</td>
<td>Q. Xiang</td>
</tr>
<tr>
<td>X22-2</td>
<td>het-6^{OR} het-c^{null} thr-2 a</td>
<td>Q. Xiang</td>
</tr>
<tr>
<td>CJ44</td>
<td>het-6^{OR} het-c^{null}; pan-2; arg-5 A</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table 3-2. List of \textit{N. crassa} strains used for heterokaryon tests.

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Relevant genotypes</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-13</td>
<td>\textit{ad3B arg-1; het-6}^{PA} \textit{het-c}^{PA} \textit{pyr-4} A</td>
<td>C. Yang</td>
</tr>
<tr>
<td>RLM58-18</td>
<td>\textit{het-6}^{PA} \textit{het-c}^{PA} \textit{pyr-4}; \textit{inl} a</td>
<td>R. L. Metzenberg</td>
</tr>
<tr>
<td>C8c158</td>
<td>\textit{ro-7} \textit{un-24} \textit{het-6}^{OR} \textit{het-c}^{PA} \textit{pyr-4} A</td>
<td>M. L. Smith</td>
</tr>
<tr>
<td>C3cr-29</td>
<td>\textit{ro-7} \textit{un-24} \textit{het-6}^{OR} \textit{het-c}^{PA} \textit{pyr-4}; \textit{inl} a</td>
<td>M. L. Smith</td>
</tr>
<tr>
<td>6-28</td>
<td>\textit{ad3B arg-1; het-6}^{PA} \textit{het-c}^{OR}; \textit{inl} A</td>
<td>C. Yang</td>
</tr>
<tr>
<td>6-19</td>
<td>\textit{ad3B arg-1; het-6}^{PA} \textit{het-c}^{OR} \textit{pyr-4}; \textit{inl} a</td>
<td>C. Yang</td>
</tr>
<tr>
<td>FGSC4571</td>
<td>\textit{ad3A; un-24} \textit{het-6}^{OR} \textit{het-c}^{OR}; \textit{nic-2; cyh-1} A^{m54}</td>
<td>A. J. F. Griffiths</td>
</tr>
</tbody>
</table>

3.2.2 Recombinant DNA techniques

Standard protocols were used for DNA manipulations (Sambrook et al., 1989). Restriction and DNA modifying enzymes were obtained from Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs and were used according to the manufacturer's instructions. PCR was performed with Perkin-Elmer Cetus DNA Thermal Cycler. The \textit{pCRII} vector (InVitrogen, San Diego, CA) was used for cloning PCR products. QIAEXII gel extraction kit (Qiagen Inc, Mississauga, Ontario) was used for recovering DNA fragments from agarose gels. Standard ligation protocols were used for DNA subcloning (Sambrook et al., 1989).
PA-type

Nc1130  LEDFPASHNYCELVLIDMEERRGHH--SPVFHVGTATKLKL----ENQFRRVRPGEYDSG-AKYAWPLVTGTFFG PA
Nc4832  LEDFPASHNYCELVLIDMEERRGHH--SPVFHVGTATKLKL----ENRQFLTRPG-HDSG-AKYVWPLVTGTFFG
Ni2316  LEDFPASHNYCELVLIDMEERRGHH--SPVFHVGTATKLKL----ENRQFLTRPG-HDSG-AKYVWPLVTGTFFG
Nt6583  LEDFPASHNYCELVLIDMEERRGHH--SPVFHVGTATKLKL----ENQFRRVRPGEYDSR-AKYAWPLVTGTFFG
Nd15923 LEDFPASHNYCELALIDHQKEDRS-KSRIFPVHTRTLKR----EGMFVPAYPEERD----YVWPLVTGTFFG
Np7221  LEDFPASHNYCELALMDIHKETE-ESQIFPVHAHRTLKL----KNGKFRRVQGERDSPPDKVWPLVTGTFFG
Gsp8243 LEDFPASHNYCELIDYKNETGR-ESQIFPVHAHTRLRY----ENGYSRVDGGGRDYSPPGKVVWPLVTGTFFG
Ss2740  LEDFPASHNYCELALMDIHKETE-ESQIFPVHAHTRLTL----KNGKFTRVQGERDDSPPDKVWPLVTGTFFG

GR-type

Nd16788 LEDFPASHNYCELVLIDHEKETCSRESIFPHVGTTRTL-NNNG-----------------KLWPLVTGTFFG
Nc1945  LEDFPASHNYCELVLIDHEKETRS-ESRIFPHVGTATRIL-NNNG-----------------KLWPLVTGTFFG GR
Nc1455  LEDFPASHNYCELVLIDHEKETRS-ESRIFPHVGTATQLT-NNNG-----------------KLWPLVTGTFFG GR
Nd1692  LEDFPASHNYCELVLIDYETENFRESQIFPVHAHTRKL-NNNG-----------------KVWPLVTGTFFG
Ss2741  LEDFPASHNYCELALIDHEKETRS-ESRIFPHVGTATRIL-NNNG-----------------KLWPLVTGTFFG
Sb7140  LEDFPASHNYCELALIYHEKETRS-ESRIFPHVGTATRIL-NNNG-----------------KLWPLVTGTFFG

OR-type

Ni3721  LEDFPASHNYCELVLIDHEKETGS-ESRIFPHVGTATRITLNRDTRNNG-----------------KVWPLVTGTFFG OR
Nc1824  LEDFPASHNYCELALIDHEKETGS-ESRIFPHVGTATRVTLLNRNTRDG-----------------KVWPLVTGTFFG OR
Nd17693 LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Sb1903  LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Sh2739  LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNE-----------------KVWPLVTGTFFG OR
Gsp8241 LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Gsp8239 LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Gsp8242 LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Ns5940  LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLLNRNTRNG-----------------KVWPLVTGTFFG OR
Ni1799  LEDFPASHNYCELVLIDMEERRGQH--SPVFHVGTNRVTLENDTRNNG-----------------KVWPLVTGTFFG OR

************ * * **** * * ************

I  II
Figure 3-1. Amino acid sequences of the specificity region from naturally occurring het-c alleles. Region underlined is variable. Asterisks indicate conserved sites. Dash represents deletion motifs. Regions I and II are polymorphic blocks. PA (Panama), OR (Oakridge), GR (Groveland) refer to the three genetically characterized alleles that confer different het-c specificity (Saupe and Glass, 1997).
3.2.3 Isolation of naturally occurring \textit{het-c} alleles

The specificity region of natural \textit{het-c} alleles was obtained by PCR amplification of genomic DNA from isolates described in Chapter 2 using Red and Yellow primers (Chapter 2, section 2.2.3). All the PCR products were cloned into \textit{pCRII} vector for further manipulations. Inferred amino acid sequences of the specificity region of the natural alleles used in this study are shown in Figure 3-1.

3.2.4 Generation of artificial \textit{het-c} alleles

Artificially constructed \textit{het-c} alleles were generated using recombinant PCR techniques (Horton et al., 1989; Vallette et al., 1989). The technique requires two sets of primers and two rounds of PCR. The principles are diagrammed in Figure 3-2 and Figure 3-3, which represent the generation of recombinant molecules and in-frame deletion/insertions, respectively. The oligonucleotides used in the construction are listed in Table 3-3. Primer sets Red and Sp1.3 were used in the second round PCR reaction for all allele constructions. Specific primer sets and templates for generating specific alleles in the first round PCR are shown in Table 3-4.

Oligonucleotide primers were synthesized by the UBC NAPS unit (Biotechnology Laboratory). All PCR reagents were supplied by Boheringer Mannheim. Plasmids containing \textit{het-c} clones were used as first round PCR templates. The first round PCR was performed by the standard protocol: initial 5 min denaturation at 94°C, followed by 30 step cycles of 1 min at 94°C, 1 min at 55°C and 1 min at 72°C, and a final extension at 72°C for 10 min. Reactions were carried out in volumes of 50 ul containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.0 mM MgCl$_2$, 25 uM each of dATP, dCTP, dGTP, dTTP, 0.2 uM of each primer, 1.25
unit of Taq polymerase and 50 ng of template plasmid DNA. The first round PCR products were purified from an agarose gel using the QIAEXII Gel Purifying kit. Two corresponding first round PCR products were combined and used as templates for the second round PCR to construct recombinant products. The second round PCR conditions were similar to that of the first round except that annealing was done at 51°C and an auto extension for 10 seconds at 72°C was performed after each cycle. PCR products from the second round were cloned into pCRII vector (Invitrogen, San Diego, CA) for further manipulations.
Figure 3-2. PCR generation of recombinant het-c alleles. The procedure is diagrammed for the construction of the pol allele that has both het-c\textsuperscript{OR} - specific and het-c\textsuperscript{PA} - specific insertion motifs; the same strategy was used to construct the rest of the recombinant het-c alleles. The oligonucleotides PO1.5 and PO1.3 were designed to have overlapping ends. The red and sp1.3 primers were used for the second round PCR for generation of all artificial het-c alleles.
Figure 3-3. PCR generation of het-c alleles by deletion. The procedure is diagrammed for the construction of pdn alleles (pd1, pd2, pd3, pd4), which have a different size of insertion in the het-c$_{PA}$ – specific motif; the same strategy was used to construct alleles del3 and od1. Allele del3 was generated from het-c$_{GR}$ with the deletion of the three residues, NNG. Allele od1 was generated from het-c$_{OR}$ with the deletion of five residues ENDTR from het-c$_{OR}$ specific motif. The oligonucleotides Pdn.5 and Pdn.3 were designed to have overlapping ends. The Red and Sp1.3 primers were used for the second round PCR for generation of all artificial het-c alleles.
Table 3-3. Oligonucleotide primers for the construction of artificial *het-c* alleles.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>GGA GAC ATG GCG ATA TCG</td>
</tr>
<tr>
<td>Sp1.3</td>
<td>GCT CAT GCC AGG AAC AAC</td>
</tr>
<tr>
<td>Del3.3</td>
<td>AGC GTA CTT GAG TTT GAG TTT AGT CGC AGT GC</td>
</tr>
<tr>
<td>Del3.5</td>
<td>ACT CAA ACT CAA GTA CGC TTG GCC CCT GGT</td>
</tr>
<tr>
<td>PO1.3</td>
<td>CTG AAC TGT CCA TTG TTA CGT GTG TCA</td>
</tr>
<tr>
<td>PO1.5</td>
<td>CAC GTA ACA ATG GAC AGT TCA GGA GCG TAA GA</td>
</tr>
<tr>
<td>PO2.3</td>
<td>GTC TAT TCT CAC GTG TGT CAT TCT TGA G</td>
</tr>
<tr>
<td>PO2.5</td>
<td>GAC ACA CGT GAG AAT AGA CAG TTC AGG CGC GTA AGA</td>
</tr>
<tr>
<td>PO3.3</td>
<td>CTG TCT ATT CTC GAG TGT CAT TCG GTG ATC</td>
</tr>
<tr>
<td>PO3.5</td>
<td>CAC ACT CGA GAA TAG ACA GTT CAG GCG CGT AAG A</td>
</tr>
<tr>
<td>PO4.3</td>
<td>GTG TCA TTC TTG AGT TTA GTC GC</td>
</tr>
<tr>
<td>PO4.5</td>
<td>CTC AAA CTC AAG AAT GAC ACA CGT AAC</td>
</tr>
<tr>
<td>P26m-3</td>
<td>CGC CCC AGA ATC ATC TCT TTC CTC ATA TGG</td>
</tr>
<tr>
<td>P26m-5</td>
<td>GAA AGA GAT GAT TCT GGG GCG AAG TAC ACG</td>
</tr>
<tr>
<td>Pd1.3</td>
<td>AGC GTA CTT TTC TTC GGG TCT TAC</td>
</tr>
<tr>
<td>Pd1.5</td>
<td>GGA GAA GGA AAG TAC GCT TGG CCC CTG</td>
</tr>
<tr>
<td>Pd2.3</td>
<td>AGC ATA CTT TAC CGC CCT GAA CTG TCT</td>
</tr>
<tr>
<td>Pd2.5</td>
<td>AGG CGC GTA AAG TAT GCT TGG CCC CTG</td>
</tr>
<tr>
<td>Pd3.3</td>
<td>AAG CGT ACT TCG GTC TTA CGC GCC TGA A</td>
</tr>
<tr>
<td>Pd3.5</td>
<td>GTA AGA CCG AAG TAC GCT TGG CCC CTG</td>
</tr>
<tr>
<td>Pd4.3</td>
<td>AGC GTA CTT TCT ATT CTC GAG TTT GAG</td>
</tr>
<tr>
<td>Pd4.5</td>
<td>CTC GAG AAT AGA AAG TAC GCT TGG CCC CTG</td>
</tr>
<tr>
<td>Od-3</td>
<td>TCC ATT GTT GAG TGT CAT TCG AGT ATC TG</td>
</tr>
<tr>
<td>Od-5</td>
<td>ATC ACA CTC AAC AAT GGA AAG TCG GTT</td>
</tr>
<tr>
<td>GO-3</td>
<td>CTT ACG TGT GTC ATT CTT GAG TGT CAT TCG AGT AGC</td>
</tr>
<tr>
<td>GO-5</td>
<td>AAG AAT GAC ACA CGT AAG TTG GTT TGG CCC TTG G</td>
</tr>
</tbody>
</table>

* The primers are shown in the 5' -to-3' orientation.
Table 3-4. Specific het-c primer sets, templates used for the 1st round of PCR to generate specific alleles.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Template</th>
<th>Primer sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>po1</td>
<td>het-c\textsuperscript{OR}</td>
<td>Red - po1.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Po1.5 - sp1.3</td>
</tr>
<tr>
<td>po2</td>
<td>het-c\textsuperscript{OR}</td>
<td>Red - po2.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Po2.5 - sp1.3</td>
</tr>
<tr>
<td>po3</td>
<td>het-c\textsuperscript{OR}</td>
<td>Red - po3.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Po3.5 - sp1.3</td>
</tr>
<tr>
<td>po4</td>
<td>het-c\textsuperscript{PA}</td>
<td>Red - po4.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{OR}</td>
<td>Po4.5 - sp1.3</td>
</tr>
<tr>
<td>po5</td>
<td>het-c\textsuperscript{OR}</td>
<td>Red - po5.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Po5.5 - sp1.3</td>
</tr>
<tr>
<td>od1</td>
<td>het-c\textsuperscript{OR}</td>
<td>Red - od1.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{OR}</td>
<td>Od1.5 - sp1.3</td>
</tr>
<tr>
<td>go1</td>
<td>het-c\textsuperscript{GR}</td>
<td>Red - go1.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{GR}</td>
<td>Go1.5 - sp1.3</td>
</tr>
<tr>
<td>del3</td>
<td>het-c\textsuperscript{GR}</td>
<td>Red - del3.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{GR}</td>
<td>Del3.5 - sp1.3</td>
</tr>
<tr>
<td>pd1</td>
<td>het-c\textsuperscript{PA}</td>
<td>Red - pd1.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Pd1.5 - sp1.3</td>
</tr>
<tr>
<td>pd2</td>
<td>het-c\textsuperscript{PA}</td>
<td>Red - pd2.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Pd2.5 - sp1.3</td>
</tr>
<tr>
<td>pd3</td>
<td>het-c\textsuperscript{PA}</td>
<td>Red - pd3.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Pd3.5 - sp1.3</td>
</tr>
<tr>
<td>pd4</td>
<td>het-c\textsuperscript{PA}</td>
<td>Red - pd4.3</td>
</tr>
<tr>
<td></td>
<td>het-c\textsuperscript{PA}</td>
<td>Pd4.5 - sp1.3</td>
</tr>
</tbody>
</table>
3.2.5 **Generation of vector cassette for construction of chimeric alleles**

Plasmids carrying $het$-$c^{OR}$ and $het$-$c^{PA}$ clones were used for generating chimeric vector cassettes. Conserved restriction sites in $het$-$c$ used for chimeric construction are shown in Figure 3-4.

The $het$-$c^{PA}$ allele was cloned as a $Xba$ I-$Eco$ R I fragment into $pCRII$ vector to give plasmid $pCRPA$ (Saupe and Glass, 1997), which has two unique restriction sites, $Stu$ I and $Sal$ I, in the specificity domain. The plasmid was then digested completely with $Stu$ I and $Sal$ I. The large fragment of approximately 6 kb was purified from agarose gel and used as a $het$-$c^{PA}$ ($Stu$ I-$Sal$ I) vector cassette that contains intact $pCRII$ vector and $het$-$c^{PA}$ with the removal of the 246 bp $Stu$ I – $Sal$ I fragment that spans the specificity region.

The $het$-$c^{OR}$ allele was cloned as a $Xba$ I-$Eco$ R I fragment in the $pCRII$ vector to give plasmid $pCROR$ that has two unique restriction sites ($Eco$ R V and $Sal$ I) in the specificity domain. The plasmid was digested completely with $Eco$ R V and $Sal$ I and the large fragment of about 6 kb was purified from agarose gel. The purified fragment was used as a $het$-$c^{OR}$ ($Eco$ R V-$Sal$ I) vector cassette that contains intact $pCRII$ vector, $het$-$c^{OR}$ with the removal of the 801 bp $Eco$ R V-$Sal$ I fragment spanning the $het$-$c$ specificity region.
Figure 3-4. Conserved restriction sites in *het-c* used for chimeric construction. The three conserved restriction sites, *EcoR* V, *Stu* I and *Sal* I were used for manipulating in-frame exchange of DNA fragments in the specificity region between *het-c*\(^{PA}\), *het-c*\(^{OR}\) and *het-c*\(^{GR}\). The *Xho* I and *Stu* I sites in *het-c*\(^{PA}\) and the *Apa* I site in *het-c*\(^{OR}\) were used for identifying chimeric constructs. The positions of PCR primers are shown in this diagram.
3.2.6 Construction of chimeric alleles

Chimeric constructs between het-c alleles were as described by Saupe and Glass (1997). The unique DNA fragment that contains the sequence encoding the 34 – 48 amino acid variable domain was exchanged in-frame between alleles. The principle is diagrammed in Figure 3-5 and applied to allelic chimeric constructions. Chimeric constructs were identified by restriction site differences within the exchanged fragment.

The specificity region of het-c<sup>OR</sup> - type alleles was digested and confirmed to contain conserved Stu I-Sal I sites. Each of the Stu I-Sal I fragments was purified from agarose gel and subcloned into het-c<sup>PA</sup>(Stu I-Sal I) vector cassette to make het-c<sup>PA(OR Stu I-Sal I)</sup> chimeric alleles.

The specificity region of het-c<sup>PA</sup> - type alleles was digested and confirmed to contain conserved EcoRV and Sal I sites. Each of the EcoRV-Sal I fragments was purified from an agarose gel and subcloned into the het-c<sup>OR</sup>(EcoRV- Sal I) vector cassette to make het-c<sup>OR(PA EcoRV Sal I)</sup> chimeric alleles. The Xho I restriction enzyme was used for identifying all the chimeric constructs.

The specificity region of GR-type alleles was digested with either Stu I-Sal I or EcoRV V-Sal I and subcloned into het-c<sup>PA</sup>(Stu I-Sal I) and het-c<sup>OR</sup>(EcoRV V-Sal I) vector cassette. Restriction enzyme Stu I was used for identifying chimeric alleles het-c<sup>PA(GR Stu I-Sal I)</sup>, and Apa I was used for identifying chimeric alleles het-c<sup>OR(GR EcoRV V-Sal I)</sup>.

All the artificial alleles have unique Stu I and Sal I sites in the specificity domain, and thus the Stu I-Sal I fragments were cloned into het-c<sup>PA(Stu I-Sal I)</sup> vector. An Xho I restriction site located within the Stu I – Sal I fragment was used for identifying chimeric constructs.
specificity region
het-c alleles

hygR

het-c

het-cOR, het-cPA, het-cGR strains
Transformation into

hyphal compartmentation and death
colon phenotype
growth rate
Figure 3-5. Principles for chimeric construction and transformation assay for *het-c* specificity. The specificity region of naturally occurring and artificially constructed *het-c* alleles was exchanged with the region of wild type *het-c* allele. The chimeric constructs were transformed into *het-c^{OR}*, *het-c^{PA}* and *het-c^{GR}* strains to assay specificity by monitoring transformant phenotype, growth rate and occurrence of hyphal compartmentation and death.
3.2.7 Cloning chimeric constructs into *N. crassa* transformation vectors

Two types of *N. crassa* transformation vectors were used: *pCB1004* with *hyg*\(^R\) as a selectable marker (Carroll et al., 1994) and *pOKE103* with *pan-2*\(^{+}\) as a selectable marker (gift of R. L. Metzenberg). Restriction maps and multiple cloning sites of both vectors are shown in Figures 3-6 and 3-7. Chimeric constructs in *pCRII* vector were digested with *EcoR* I and *Xba* I and then subcloned into both *pCB1004* and *pOKE103* vectors.

3.2.8 Construction of strain CJ44

A *het-c* deletion strain *X22-2* (*het-6*\(^{OR}\) *het-c*\(^{null}\) *thr-2 a*) was isolated from strain C9-2 (*het-6*\(^{OR}\) *het-c*\(^{PA}\) *thr-2 a*; Xiang, unpublished result). To obtain *het-c* null strain CJ44 (*het-6*\(^{PA}\) *het-c*\(^{null}\); *pan-2*; *arg-5 a*), strain *Xa-3* (*het-6*\(^{OR}\) *het-c*\(^{PA}\); *pan-2*; *arg-5 A*; Xiang unpublished) was used as a female and crossed with strain *X22-2*. Ascospores produced in the cross were collected from Petri dish lids, streaked on agar plate with the supplements pantothenic acid (PAN) and arginine (ARG), and heat-shocked in a 65°C oven for 45 min. Plates were cooled to room temperature and then incubated overnight at 30°C. Germinated ascospores were picked from plates under a microscope and transferred to 1 ml of Vogel's agar media (Vogel, 1964) plus PAN and ARG. Progeny were first screened for PAN and ARG requirements. The *pan-2*, *arg-5* progeny were tested for mating type on a lawn of *fl A* (FGSC4317) and *fl a* (FGSC4347) tester strains, followed by *het-6* heterokaryon test with tester strains (Table 3-2). Six progeny that were *het-6*\(^{OR}\); *pan-2*; *arg-5 A* were selected and used in *het-c* heterokaryon tests with *het-c* tester strains (Table 3-2). Strain CJ44 displayed dual compatibility with both *het-c*\(^{PA}\) and *het-c*\(^{OR}\) tester strains and thus was designated as a *het-c*\(^{null}\) strain.
Figure 3-6. Plasmid map of pCB1004 (Carroll et al., 1994).
Figure 3-7. Plasmid map of $pOKE103$ (gift of R. L. Metzenberg).

$pOKE103$

6500 bp
3.2.9 DNA transformation

The specificity of the chimeric constructs was tested by transformation assays in *N. crassa*. For *pCB1004* as a transformation vector, C9-2 (*het-6^{OR} het-c^{PA} thr-2 a*), C2-2-9 (*het-6^{PA} het-c^{OR} thr-2 A*) and FGSC 2193 (*het-6^{OR} het-c^{GR} A*) were used as recipient strains (Table 3-1). For *pOKE103* as a transformation vector, C15-1a (*het-6^{OR} het-c^{OR}; pan-2; arg-5; inl a*) and Xa-3 (*het-6^{OR} het-c^{PA}; pan-2; arg-5 A*) were used as recipient strains (Table 3-1).

Spheroplasts were prepared according to Schweizer et al. (1981) and stored at $-70^\circ$C until use.

A modified procedure of the *P. anserina* transformation protocol was applied to *N. crassa* (S. J. Saupe, personal communication). Fifty ml of spheroplasts were thawed on ice and then heat shocked at $48^\circ$C for 5 minute, followed by a 30 second incubation on ice. The spheroplasts were placed at room temperature for 10 minute before DNA was added. One µg of DNA was used for each 50 µl of spheroplasts. The mixture of spheroplasts and DNA was incubated at room temperature for an additional 10 minute and then added to 1 ml of 40% PEG, 10 mM MOPS, 50 mM CaCl$_2$. The mixture was further incubated at room temperature for 10 minute. Seven ml of pre-warmed top agar was added into the mixture and poured into bottom agar plates. Transformation plates were incubated at 30$^\circ$C for three days. A population of fifteen colonies was transferred to individual plates to assay vegetative incompatibility by monitoring colony morphology, growth rate and occurrence of hyphal compartmentation and death.

For co-transformation, 0.5 µg DNA each of *pOKE103* construct and *pCB1004* construct was used. Strain CJ44 (*het-6^{PA} het-c^{N}; pan-2; arg-5 a*) was used as a co-transformation recipient (Table 3-1).
3.2.10 **Heterokaryon incompatibility tests**

Heterokaryon incompatibility tests were performed by spotting conidial suspensions of auxotrophic strains onto Vogel’s minimal agar medium (Vogel, 1964). Strains forming vigorous conidiating cultures after three days incubation at 30°C were considered compatible. Relative genotype and origin of the tester strains are listed in Table 3-2.

3.2.11 **Growth rate measurements**

Linear growth rate of transformants was measured in race tubes as described by Jacobson et al. (1998). A colony cut from a transformation plate was placed at one end of a 40 ml glass tube containing 25 ml of medium. All tubes were incubated at 25°C. The starting point was marked as the leading edge of the colony after overnight growth; subsequent growth was recorded as the distance to the leading edge of the colony at ~24 hr intervals. Each experiment was performed in three replicates with different colonies from the same transformation experiment.

3.2.12 **Light microscopy**

The stain Evan’s Blue (Direct Blue 53, CI23860, Aldrich Chemical Co., Milwaukee, Wisconsin) is excluded from cells with intact plasma membranes (Gaff and Okong’O-Ogola, 1971) and was thus used to identify dead and dying hyphal compartments. A transformant was inoculated onto a cellophane membrane layered on top of agar plates and incubated at 30°C. Cellophane membranes with mycelium adhering were removed from the medium, placed on glass slides and flooded with 1% Evans blue in water (w/v). After 5 min, mycelia were rinsed in distilled water, floated off the cellophane and onto a glass slide, and mounted with 90% glycerol in 0.1 M phosphate buffer (pH 7.2) under a cover slip. Samples were
examined under bright field illumination on a Olympus microscope and photographed on Kodak TMAX-400 film.
3.3 RESULTS

An exchange of an approximately 220 bp (amino acid position 223 - 294) Stu-I-Sal I fragment in the variable domain between het-c alleles was shown to be sufficient to result in a switch of allelic specificity (Saupe and Glass, 1997). This principle was used for chimeric construction in this study to investigate the het-c specificity of natural and artificial mutant het-c alleles. Chimeric constructs were introduced into het-c\(^OR\), het-c\(^PA\) and het-c\(^GR\) strains, and het-c specificity was monitored by colony morphology, linear growth rate and occurrence of hyphal compartmentation and death.

3.3.1 Phenotypes of transformants with standard het-c alleles

Typically, transformants with standard het-c alleles in N. crassa exhibit three classes of phenotypes that are distinguished by the morphology of transformants, linear growth rate and occurrence of hyphal compartmentation and death (Table 3-5). The colony morphology of the three classes of transformants is shown in Figure 3-8.

Class I, compatible transformants (++) such as introduction of a het-c allele (e. g. het-c\(^PA\)) into a recipient strain with the same het-c background (e. g. strain C9-2 het-c\(^PA\) in Table 3-1), display vigorous growth and normal conidiation (Figure 3-8) that is indistinguishable from wild-type strains. However, the growth rate varies among different recipient strains. Compatible transformants in the het-c\(^PA\) recipient strain have a growth rate of 6.0 cm / day; compatible transformants in either het-c\(^GR\) or het-c\(^OR\) recipient strains have a growth rate of 4.9 – 5.2 cm /day (Figure 3-9). The growth rate differences between the class I transformants may be due to the background difference of the recipient strains and are not
statistically significant (t test, \( \alpha < 0.05 \)). The only unexpected member of the class I transformants was the \( \text{het-c}^{GR} \) strain containing the \( \text{het-c}^{PA} \) construct; the compatible phenotype in this case may be due to a modifier in the \( \text{het-c}^{GR} \) strain (S. J. Saupe, personal communication; see discussion 3.4.1). The occurrence of hyphal compartmentation and death in these transformants was examined by Evan’s blue staining. No significant occurrence (< 1%) of hyphal compartmentation and death was observed and the hyphae exhibit morphology of a wild type strain (Figure 3-10).

Table 3-5. Three classes of phenotypes of \( \text{het-c} \) transformants

<table>
<thead>
<tr>
<th>Rating class</th>
<th>Occurrence of death, hyphal compartments</th>
<th>Linear growth rate (cm / day)</th>
<th>Colony morphology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I (+++)</td>
<td>No hyphal death, wild type hyphal morphology</td>
<td>&gt; 4.0</td>
<td>Conidiation. Vigorous growth.</td>
</tr>
<tr>
<td>Class II (-)</td>
<td>Budding-like hyphal morphology, hyphal death occurred after 2 days.</td>
<td>1.0 – 2.5</td>
<td>Little conidiation. Poor growth.</td>
</tr>
<tr>
<td>Class III (-)</td>
<td>Hyphal death occurred within 1 day.</td>
<td>&lt; 1.0</td>
<td>No conidia (aconidial). Curly, severely inhibited hyphal growth.</td>
</tr>
</tbody>
</table>

The class II transformants displayed an intermediate incompatible phenotype (-). The introduction of either \( \text{het-c}^{PA} \) or \( \text{het-c}^{GR} \) into a strain with \( \text{het-c}^{OR} \) background (e.g., \( C2-2-9 \) in Table 3-1) or the introduction of \( \text{het-c}^{GR} \) into a strain with \( \text{het-c}^{PA} \) background (e.g., \( C9-2 \) in Table 3-1) gave this class of transformants. They showed a linear growth rate of 1 - 2.5 cm per day (Figure 3-9), scarce conidiation (Figure 3-8), budding-like hyphal morphology (Figure 3-10) and hyphal compartmentation and death that occur after two days of growth.
The class III type (--) of transformants show a severely incompatible phenotype characterized by slow-growth (linear rate < 1.0 cm / day; Figure 3-9), curly, flat aconidial morphology (Figure 3-8), and hyphal death occurring in >20% of the hyphal segments within one day of growth (Figure 3-10). This class of transformants is represented by introduction of \textit{het-c}^{OR} into a recipient strain of either \textit{het-c}^{PA} or \textit{het-c}^{GR} background.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3-8.png}
\caption{Colony morphology of introduction wild-type \textit{het-c} alleles (PA, \textit{het-c}^{PA}; OR, \textit{het-c}^{OR}; GR, \textit{het-c}^{GR}) into a \textit{het-c}^{OR} (C2-2-9, C15-1), \textit{het-c}^{PA} (C9-2, Xa-3) and \textit{het-c}^{GR} (FGSC2193) background.}
\end{figure}
Recipient strains

<table>
<thead>
<tr>
<th>Alleles</th>
<th>into PA</th>
<th>into OR</th>
<th>into GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3-9.** Average growth rate (cm/day) of standard (wild-type) *het-c* transformants. Experiments were performed in triplicates. Error bar indicates standard deviation. Geneotypes of recipient strains: PA, \(het-c^PA\); OR, \(het-c^{OR}\); GR, \(het-c^{GR}\).

**Figure 3-10.** Hyphal morphology and occurrence of hyphal compartmentation and death in the three classes of *het-c* transformants after 1 day and 3 days of growth (Table 3-5). Hyphal compartmentation and death were visualized by staining with Evan’s blue for 5 min (See Materials and Methods, section 3.2.11). Magnification 72X.
3.3.2 Amino acid sequence variability of het-c alleles

To determine whether the amino acid sequences or the pattern of insertion/deletion in the variable domain is most important for het-c allelic specificity, I took advantage of the amino acid sequence diversity in the variable domain of naturally occurring het-c alleles (see Chapter 2, Figure 2-4). The specificity of the natural alleles was assayed by the introduction of chimeric constructs into recipient strains of different het-c backgrounds. Twenty-four of the natural alleles (Figure 3-1), representing amino acid substitutions at multiple codon positions in the variable region and retaining the conserved Stu I, Sal I and EcoR V restriction sites, were selected for chimeric allele construction.

The het-c alleles clustered into three het-c allelic types by their pattern of insertions/deletions as previously described (see Chapter 2; and Saupe and Glass, 1997). Two regions in the het-c variable domain that show amino acid differences among isolates are denoted as region I and II (Figure 3-1). The 24 isolates fall into two groups based on amino acid variations within region I. The first group contains a M(G)EERRGG(Q)H consensus sequence within region I. The second group of alleles is much more variable in amino acid sequence in region I, but has a consensus of IH(Y)E(Q/K)K(N)ET(N/D)G(R/P/C)S(E/R). Less amino acid variability is observed among the 24 isolates in region II, and in particular a FPHVG motif is completely conserved in this region among all alleles. Chimeric construction and specificity of alleles in the three allelic groups are described in detail below.
3.3.3 Chimeric construction and specificity of \textit{het-c}^{OR} – type of alleles

The amino acid sequence variability in the specificity region of the ten OR-type alleles is shown in Figure 3-1. Two types of consensus sequences are observed in the variable region I, IHEKETGS and M(G)EERRGQH. Amino acids in region II vary at multiple sites.

In order to distinguish chimeric constructs, the 220 bp of the \textit{Stu I} – \textit{Sal I} fragment of the specificity region of all ten OR-type alleles was cloned into the \textit{het-c}^{PA} (\textit{Stu I} – \textit{Sal I}) vector cassette. Chimeric \textit{het-c}^{PA} (\textit{OR Stu I} – \textit{Sal I}) constructs were identified by the absence of an \textit{Xho I} restriction site, which is present in \textit{het-c}^{PA} and absent in the \textit{het-c}^{OR} specificity region (data not shown).

Each chimeric construct was introduced into \textit{het-c}^{OR}, \textit{het-c}^{PA} and \textit{het-c}^{GR} strains (Table 3-1) and the phenotype of the transformants was examined compared to transformants carrying wild-type \textit{het-c} alleles. Regardless of the amino acid sequence variability in region I (for example, alleles \textit{Ni3721} and \textit{Ndi6793}; Figure 3-1) and region II, all the OR-type alleles showed the same \textit{het-c} specificity. All the transformants displayed class I phenotype (compatibility) when each chimeric construct was introduced into a \textit{het-c}^{OR} strain. All gave class III (severe incompatibility) transformants when introduced into either \textit{het-c}^{PA} or \textit{het-c}^{GR} strains (Table 3-6).

All the class I type of transformants had an average linear growth rates of 4.3 – 5.7 cm per day (Table 3-7), and displayed vigorous growth and normal conidiation phenotypes (representative phenotype is shown in Figure 3-11). Treatment of hyphae in the class I transformants with Evan’s blue did not show an occurrence of a significant level of dead hyphal compartments (data not shown).
All the class III type of transformants displayed severely incompatible phenotypes and had linear growth rates of $< 0.5$ cm per day (Table 3-7). These transformants showed a flat, curling, aconidial morphology (representative phenotype is shown in Figure 3-11). The occurrence of dead hyphal compartments could be detected within one day of vegetative growth (data not shown). The number of dead hyphal compartments reached approximately 30% of the population after two days of growth.

**Table 3-6.** Summary of the phenotypes of transformants containing $het-c^{OR}$ - type chimeric constructs into *N. crassa* strains.

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Phenotypes of transformants in recipient strains of different $het-c$ specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$het-c^{FA}$</td>
</tr>
<tr>
<td>$Gsp8241$</td>
<td>--</td>
</tr>
<tr>
<td>$Gsp8242$</td>
<td>--</td>
</tr>
<tr>
<td>$Nc1824$</td>
<td>--</td>
</tr>
<tr>
<td>$Nc847$</td>
<td>--</td>
</tr>
<tr>
<td>$Ns5940$</td>
<td>--</td>
</tr>
<tr>
<td>$Ns963$</td>
<td>--</td>
</tr>
<tr>
<td>$Ndi6793$</td>
<td>--</td>
</tr>
<tr>
<td>$Sh2739$</td>
<td>--</td>
</tr>
<tr>
<td>$Sb1903$</td>
<td>--</td>
</tr>
<tr>
<td>$Ni1940$</td>
<td>--</td>
</tr>
<tr>
<td>$Ni3721$</td>
<td>--</td>
</tr>
<tr>
<td>$Ni1799$</td>
<td>--</td>
</tr>
<tr>
<td>$Gsp8239$</td>
<td>--</td>
</tr>
</tbody>
</table>

Classification of transformants was described in Table 3-5.
Table 3-7. Average rate of growth (cm / day) (mean ± SD) of the transformants containing naturally occurring \textit{het-c}^{OR} - type chimeric constructs. All experiments were done in triplicates.

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Average growth rate of transformants in recipient strains of different \textit{het-c} specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>\textit{het-c}^{PA}</td>
</tr>
<tr>
<td>\textit{Gsp8241}</td>
<td>0.2 ± 0.05</td>
</tr>
<tr>
<td>\textit{Gsp8242}</td>
<td>0.2 ± 0.00</td>
</tr>
<tr>
<td>\textit{Nc1824}</td>
<td>0.3 ± 0.10</td>
</tr>
<tr>
<td>\textit{Nc847}</td>
<td>0.2 ± 0.05</td>
</tr>
<tr>
<td>\textit{Ns5940}</td>
<td>0.2 ± 0.00</td>
</tr>
<tr>
<td>\textit{Ns963}</td>
<td>0.4 ± 0.13</td>
</tr>
<tr>
<td>\textit{Nd16973}</td>
<td>0.3 ± 0.12</td>
</tr>
<tr>
<td>\textit{Sh2739}</td>
<td>0.3 ± 0.13</td>
</tr>
<tr>
<td>\textit{Sb1903}</td>
<td>0.3 ± 0.11</td>
</tr>
<tr>
<td>\textit{Ni1940}</td>
<td>0.3 ± 0.10</td>
</tr>
<tr>
<td>\textit{Ni3721}</td>
<td>0.2 ± 0.00</td>
</tr>
<tr>
<td>\textit{Ni1799}</td>
<td>0.2 ± 0.00</td>
</tr>
<tr>
<td>\textit{Gsp8239}</td>
<td>0.3 ± 0.12</td>
</tr>
</tbody>
</table>

Figure 3-11. Representative colony morphology of transformants containing \textit{het-c}^{OR} - type chimeric constructs. a, b, c, introduction of \textit{het-c}^{OR} chimeric constructs into recipient \textit{N. crassa} strain \textit{het-c}^{OR} (C2-2-9), \textit{het-c}^{GR} (FGSC2193) and \textit{het-c}^{PA} (C9-2), respectively.
3.3.4 Chimeric construction and specificity of het-c^{GR}-type alleles

The specificity region of GR-type alleles lacks both the OR-specific and PA-specific insertions (Figure 3-1). Restriction enzyme digestion and sequence analysis (data not shown) indicated that the specificity regions of all the GR-type alleles lack both *Apa* I and *Xho* I sites that are present in the OR-specific and PA-specific insertions, respectively. These alleles all contain a similar motif in region I, although amino acid variability and a single amino acid insertion/deletion is observed in this region. In region II, amino acid variability occurs at one or two amino acid sites (i.e., TA/VTR/QITL) in all GR-type alleles with the exception of *Ndo1692* (AHTRKTL) that has three to four amino acid residue differences from remaining of the GR-type alleles (Figure 3-1).

The *EcoR* V- *Sal* I or *Stu I* - *Sal* I fragments of the specificity region from the six GR-type alleles (Figure 3-1) were cloned into either the het-c^{OR}(*EcoR* V - *Sal* I) or het-c^{PA}(*Stu I* - *Sal* I) vector cassette. The het-c^{OR} (*GR* *EcoR* V - *Sal* I) chimeric constructs were confirmed by the absence of an *Apa* I site located within the OR-specific insertion region. The het-c^{PA} (*GR* *Stu I* - *Sal* I) chimeric constructs were distinguished from het-c^{PA} allele by the absence of the *Xho* I site located within the PA-specific insertion region.

Both het-c^{PA} (*GR* *Stu I* - *Sal* I) and het-c^{OR} (*GR* *EcoR* V - *Sal* I) chimeric constructs gave class I transformants when introduced into het-c^{GR} spheroplasts and gave class II transformants when introduced into either het-c^{OR} or het-c^{PA} strains (Table 3-8). All the class I type transformants had vigorous growth with average growth rates of 4.0 – 4.8 cm per day (Table 3-9). No significant (less than 1%) hyphal compartmentation or death was observed. All the class II transformants displayed intermediate incompatibility phenotype and had an average linear growth rates of 1.4 - 2.2 cm per day (Table 3-9) and scarce conidiation. Microscopic
examination of these transformants revealed abnormal budding-like hyphal morphology and the presence of dead hyphal compartments after approximately 48 hours of growth (data not shown).

**Table 3-8.** Transformation of chimeric constructs of naturally occurring mutant \( het-c^{GR} \)-type alleles into *N. crassa*. Classification of transformants was described in Table 3-5.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Phenotype of transformants in recipient strains of different ( het-c ) specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( het-c^{PA} )</td>
</tr>
<tr>
<td><em>Nc1945</em></td>
<td>-</td>
</tr>
<tr>
<td><em>Sh7140</em></td>
<td>-</td>
</tr>
<tr>
<td><em>Ss2741</em></td>
<td>-</td>
</tr>
<tr>
<td><em>Nc1455</em></td>
<td>-</td>
</tr>
<tr>
<td><em>Ndo1692</em></td>
<td>-</td>
</tr>
<tr>
<td><em>Ndi6788</em></td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3-9. Average growth rate (cm / day) (mean ± SD) of transformants containing \( \text{het-c}^{GR} \) chimeric constructs. All experiments were done in triplicates.

<table>
<thead>
<tr>
<th>Allele</th>
<th>( \text{het-c}^{PA} ) (Average growth rate)</th>
<th>( \text{het-c}^{OR} ) (Average growth rate)</th>
<th>( \text{het-c}^{GR} ) (Average growth rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Ncl945} )</td>
<td>1.6 ± 0.15</td>
<td>1.4 ± 0.10</td>
<td>4.6 ± 0.10</td>
</tr>
<tr>
<td>( \text{Sh7140} )</td>
<td>1.4 ± 0.10</td>
<td>2.2 ± 0.30</td>
<td>4.5 ± 0.12</td>
</tr>
<tr>
<td>( \text{Ss2741} )</td>
<td>1.7 ± 0.23</td>
<td>1.8 ± 0.20</td>
<td>4.7 ± 0.15</td>
</tr>
<tr>
<td>( \text{Nc1455} )</td>
<td>1.5 ± 0.15</td>
<td>2.0 ± 0.24</td>
<td>4.5 ± 0.12</td>
</tr>
<tr>
<td>( \text{Ndo1692} )</td>
<td>1.8 ± 0.17</td>
<td>2.3 ± 0.30</td>
<td>4.0 ± 0.10</td>
</tr>
<tr>
<td>( \text{Ndi6788} )</td>
<td>1.7 ± 0.18</td>
<td>1.9 ± 0.18</td>
<td>4.8 ± 0.20</td>
</tr>
</tbody>
</table>

3.3.5 Chimeric construction and specificity of \( \text{het-c}^{PA} \) – type alleles

The PA-type alleles fall into either one of two groups in region I (Figure 3-1) and also display a high degree of amino acid variability within region II. Differences in the specificity region of PA-type alleles as compared to the GR-type and OR-type of alleles include (Figure 3-1): 1) they contain a large insertion; 2) they show a high degree of variability in amino acid composition in the insertion region among the eight alleles; 3) an exceptional allele containing an insertion of a different length, \( \text{Ndi5923} \), was identified.

Due to the additional \( \text{Stu} \) I site in the PA-specific insertion region in all PA-type alleles as compared to OR-type or GR-type alleles, the \( \text{EcoR V-Sal I} \) fragment of the specificity region of the PA-type alleles was used to replace the same fragment from \( \text{het-c}^{OR} \) for chimeric construction. The unique \( \text{Xho I} \) restriction site located within the PA-specific
insertion region was used to distinguish chimeric constructs of $\text{het-c}^{\text{OR}(\text{PA EcoRV} \rightarrow \text{SalI})}$ from $\text{het-c}^{\text{OR}}$ alleles (data not shown).

The specificity of naturally occurring PA-type alleles is summarized in Table 3-10. With the one exception of $\text{Ndi5923}$ (named allele $p26$ later), all other $\text{het-c}^{\text{PA}}$ chimeric alleles gave class I transformants when introduced into the $\text{het-c}^{\text{PA}}$ and $\text{het-c}^{\text{GR}}$ strains, and gave class II transformants when introduced into $\text{het-c}^{\text{OR}}$ strains. The class I (compatible) transformants showed an average growth rate of $5.5 - 6.5$ cm per day in the $\text{het-c}^{\text{PA}}$ recipient strain and of $4.5 - 5.2$ cm per day in the $\text{het-c}^{\text{GR}}$ recipient strain (Table 3-11). The growth rate differences of the class I transformants in different recipient strains are consistent with the observations of transformants containing standard $\text{het-c}$ alleles (see section 3.3.1). The class II transformants showed an average growth rate of $1.4 - 1.9$ cm per day (Table 3-11). Dead hyphal compartments were not observed in the class I transformants in either $\text{het-c}^{\text{PA}}$ or $\text{het-c}^{\text{GR}}$ recipient strains but occurred in the class II transformants after two days of growth (data not shown).

The allele $\text{Ndi5923}$ has a PA-specific pattern of insertion but is shorter (four to six codons shorter, depending on the reference allele). Unlike the other PA-type of chimeric constructs, the $\text{Ndi5923}$ construct gave class II transformants when introduced into $\text{het-c}^{\text{PA}}$, $\text{het-c}^{\text{OR}}$ and $\text{het-c}^{\text{GR}}$ strains. The class II transformants in the three recipient strains display a similar phenotype, characterized by linear growth rate at approximately $1.3 - 1.8$ cm per day (Table 3-11), abnormal budding-like hyphal morphology and the occurrence of dead hyphal compartments after approximately two days of growth (Figure 3-12). This observation indicates that allele $\text{Ndi5923}$ ($p26$) has a $\text{het-c}$ specificity different from that of $\text{het-c}^{\text{OR}}$, $\text{het-c}^{\text{PA}}$ or $\text{het-c}^{\text{GR}}$ - type alleles.
It is possible that the class II phenotype of p26 transformants was due to self-incompatibility of the allele p26. To exam this possibility, chimeric constructs of p26 allele were introduced into a het-c null strain X22-2 (het-6°R het-c null thr-2 a). Only class I transformants were observed. These data indicated that the incompatible phenotype of the p26 chimeric transformants was not due to the self-incompatibility of the p26 allele but due to a modification of het-c specificity. Heterokaryons between the X22-2 transformants containing the p26 allele with the het-c tester C8c158 (het-6°R het-c PA pyr-4 A) and FGSC4571 (het-6°R het-c OR pyr-4; ad3B A m54) (Table 3-2) displayed a class II incompatible phenotype (data not shown), which is consistent with the observation that p26 allele confers a novel het-c specificity.
Table 3-10. Transformation of chimeric constructs of naturally occurring $het-c^{PA}$ - type alleles into $N. crassa$.

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Phenotype of transformants in recipient strains of different $het-c$ specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$het-c^{PA}$</td>
</tr>
<tr>
<td>$Nc1130$</td>
<td>++</td>
</tr>
<tr>
<td>$Nc4711$</td>
<td>++</td>
</tr>
<tr>
<td>$Nc4832$</td>
<td>++</td>
</tr>
<tr>
<td>$Ni2316$</td>
<td>++</td>
</tr>
<tr>
<td>$Nt6583$</td>
<td>++</td>
</tr>
<tr>
<td>$Nd3228$</td>
<td>++</td>
</tr>
<tr>
<td>$Nd5923$</td>
<td>-</td>
</tr>
<tr>
<td>$Np7221$</td>
<td>++</td>
</tr>
<tr>
<td>$Ss2740$</td>
<td>++</td>
</tr>
<tr>
<td>$Gsp8243$</td>
<td>++</td>
</tr>
</tbody>
</table>

Classification of transformants is described in Table 3-5. All the $het-c^{PA}$ specificity type alleles gave compatible transformants when introduced into $het-c^{GR}$ strain. It may be due to a modifier in the $het-c^{GR}$ strain (see discussion 3.4.1). $Nd5923$ has a different insertion size in the PA-specific motif and demonstrated incompatibility with strains containing any of the three $het-c$ allelic types.
Table 3-11. Average rate of growth (cm / day) (mean ± SD) of transformants containing \textit{het-c}^{PA} chimeric constructs. All experiments were done in triplicates.

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Average growth rate of transformants in recipient strains of different \textit{het-c} specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>\textit{het-c}^{PA}</td>
</tr>
<tr>
<td>\textit{Nc1130}</td>
<td>6.3 ± 0.15</td>
</tr>
<tr>
<td>\textit{Nc4832}</td>
<td>5.5 ± 0.30</td>
</tr>
<tr>
<td>\textit{Ni2316}</td>
<td>6.0 ± 0.22</td>
</tr>
<tr>
<td>\textit{Nt6583}</td>
<td>6.2 ± 0.20</td>
</tr>
<tr>
<td>\textit{Ndi3228}</td>
<td>5.9 ± 0.28</td>
</tr>
<tr>
<td>\textit{Ndi5923}</td>
<td>1.8 ± 0.10</td>
</tr>
<tr>
<td>\textit{Np7221}</td>
<td>6.2 ± 0.20</td>
</tr>
<tr>
<td>\textit{Ss2740}</td>
<td>6.5 ± 0.35</td>
</tr>
<tr>
<td>\textit{Gsp8243}</td>
<td>6.0 ± 0.10</td>
</tr>
</tbody>
</table>
Figure 3-12. Hyphal compartmentation and death of Ndi5923 (p26) chimeric construct introduced into different het-c strains after one day and three days of growth. Dead hyphal compartments were observed by staining with Evan's blue for 5 min. a, p26 construct introduced into het-cOR strain. b, p26 construct introduced into het-cGR strain. c, p26 construct introduced into het-cPA strain. Magnification 64X.

3.3.6 Summary of het-c specificity of naturally occurring alleles

Chimeric construction and transformation assays indicated that all the OR-type and GR-type of alleles confer the same specificity as identified in the standard het-cOR and het-cGR alleles (Nc1824 and Nc1945; Saupe and Glass, 1997), respectively. With the exception of Ndi5923, all other PA-type alleles confer the same specificity as the standard het-cPA allele (Nc1130; Saupe and Glass, 1997). These data indicate that the naturally occurring variability in region I and region II (Figure 3-1) does not affect het-c specificity. However, the allele Ndi5923, which has a PA-type insertion of a different size, confers a novel het-c specificity. These data suggest that a new het-c specificity could be gained by altering the size of the PA-type insertion or by altering the amino acid composition of the PA-specific insertion.
3.3.7 Artificial PA-type alleles with size variation in the PA-specific insertion

As described in the above section, allele p26 \((Ndi5923)\) gave class II transformants when introduced into \(het-c^{PA}\), \(het-c^{OR}\) and \(het-c^{GR}\) recipient strains. It is possible that incompatibility is due to amino acid variation in the sequences in the PA-specific insertion motif. To examine this possibility, an artificial allele \(p26m\) derived from \(p26\) by the addition of five amino acids, DSGAK, was constructed (Figure 3-13). The allele \(p26m\) conferred the same specificity as a \(het-c^{PA}\) allele \((Ncl1130)\). The \(p26m\) chimeric construct gave class I (compatible) transformants when introduced into \(het-c^{PA}\) and \(het-c^{GR}\) strains and gave class II (intermediate incompatible) transformants when introduced into \(het-c^{OR}\) spheroplasts (Figure 3-14). The class I transformants showed vigorous growth and a growth rate of 5.7 cm per day in \(het-c^{PA}\) strain and 4.8 cm per day in \(het-c^{GR}\) strain (Figure 3-15). The class II transformants showed inhibited growth and a growth rate of 1.4 cm per day (Figure 3-15).

\[
\begin{aligned}
\text{PA} & \quad \text{VLIDMEERRGGH-SPVFPHVGTATKLKLENQFRRVPGEYDSGAKYAWPLVT} \\
\text{P26} & \quad \text{ALIDIHQKEDRSKSRIFPHVGTRTKLKERNGMFVPAPYEERV-D-----YVWPLVT} \\
\text{P26m} & \quad \text{ALIDIHQKEDRSKSRIFPHVGTRTKLKERNGMFVPAPYEERD\textbf{DSGAK}YVWPLVT} \\
\text{*} & \quad \text{**} & \quad \text{***} & \quad \text{****} & \quad \text{*****} & \quad \text{******}
\end{aligned}
\]

**Figure 3-13.** Amino acid sequences comparison of \(HET-c^{PA}\) \((Ncl1130)\), P26, and P26m specificity region. Note that P26m has an addition of five amino acids, DSGAK. Asterisks indicate identical sites.
Figure 3-14. Phenotype comparison of $het-c^{PA}$, p26m, p26 transformants after three days of vegetative growth. Genotypes of recipient strains are shown in Table 3-1. PA, $het-c^{PA}$; OR, $het-c^{OR}$; GR, $het-c^{GR}$.

Figure 3-15. Average growth rate comparison of $het-c^{PA}$, $het-c^{OR}$ and $het-c^{GR}$ transformants containing either $het-c^{PA}$, p26 or p26m chimeric constructs. Bar indicates standard deviation. All experiments were done in triplicates. Genotypes of recipient strains: PA, $het-c^{PA}$ (C9-2); OR, $het-c^{OR}$ (C2-2-9); GR, $het-c^{GR}$ (FGSC 2193).
The data from the naturally occurring het-c alleles suggested that the size of the insertion or the added amino acids affected het-c allele specificity. To investigate how the insertion size affects allele specificity, a set of artificial het-c alleles, pd1, pd2, pd3 and pd4, were generated from het-c\textsuperscript{PA} that contained different numbers of amino acids in the PA-specific insertion region (Figure 3-16). The pd1 allele has a deletion of five amino acids (A, G, S, D, Y) and has the same size of insertion in the specificity region as p26. The pd2 allele has a deletion of seven residues (A, G, S, D, Y, G, E). The pd3 allele has a deletion of ten amino acids (A, G, S, D, Y, G, E, G, P, R). The pd4 allele has the complete deletion of the fifteen amino acids in the PA-specific insertion motif. The composition of alleles pd1, pd2 and pd3 was confirmed by DNA sequencing (see Appendix 7.2). Allele pd4 was distinguished from het-c\textsuperscript{PA} by the Stu I restriction site within the het-c\textsuperscript{PA} - specific insertion region (data not shown).

The phenotypes of the transformants containing the four artificial het-c\textsuperscript{PA} alleles are summarized in Table 3-12. I describe the characteristics of vegetative growth of the transformants in the following paragraphs.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>VLIDMEERRGHH-SPVFPHVGTATKLKL---ENRQFRRVRPGEYDSGAKYAWPLVT</td>
</tr>
<tr>
<td>GR</td>
<td>VLIDIHEKETRSESRIPPHVGTATRITL----NNG---------------KLVWPLVT</td>
</tr>
<tr>
<td>OR</td>
<td>VLIDMEERRGHH-SPVFPHVGTDTVRTLRNDTRNNG---------------KSVWPLVT</td>
</tr>
<tr>
<td>P26</td>
<td>ALIDIHQKEDRSKSRIFPHVGTTRTKLR----ENGMFVPAPYEERD-----YVWPLVT</td>
</tr>
<tr>
<td>PD1</td>
<td>VLIDMEERRGHH-SPVFPHVGTATKLKL---ENRQFRRVRPGEYDSGAKYAWPLVT</td>
</tr>
<tr>
<td>PD2</td>
<td>VLIDMEERRGHH-SPVFPHVGTATKLKL---ENRQFRRVRPGEYDSGAKYAWPLVT</td>
</tr>
<tr>
<td>PD3</td>
<td>VLIDMEERRGHH-SPVFPHVGTATKLKL---ENRQFRRVRPGEYDSGAKYAWPLVT</td>
</tr>
<tr>
<td>PD4</td>
<td>VLIDMEERRGHH-SPVFPHVGTATKLKL---ENRQFRRVRPGEYDSGAKYAWPLVT</td>
</tr>
</tbody>
</table>

Figure 3-16. Amino acid sequence comparison of HET-c\textsuperscript{PA}, HET-c\textsuperscript{GR}, HET-c\textsuperscript{OR}, P26, PD1, PD2, PD3 and PD4 within the variable region.
When chimeric constructs were introduced into the $het-c^{GR}$ strain (FGSC2193), $pd1$, $pd2$, and $pd3$ gave class II transformants (Table 3-12). These class II transformants showed an average growth rate of 1.7 – 2.0 cm per day (Figure 3-17). The colonies displayed flat growth and little conidiation (Figure 3-18). Dead hyphal compartments occurred after 3 days of vegetative growth (data not shown). Introduction of allele $pd4$ that has a GR-type deletion pattern into $het-c^{GR}$ spheroplasts gave class I (compatible) transformants (Table 3-12).

All four alleles gave class II transformants when introduced into a $het-c^{PA}$ strain (C9-2 or Xa-3). These transformants displayed the average growth rate of 1.4 – 1.7 cm per day (Figure 3-17) with sparse conidiation (Figure 3-18). When hyphae from these transformants were stained with Evan’s Blue, dead hyphal compartments were observed at three days of growth and hyphae showed an abnormal, budding-like morphology (Figure 3-19).

Introduction of $pd1$, $pd2$, $pd3$ and $pd4$ chimeric constructs into a $het-c^{OR}$ strain (C2-2-9 or C15-1) gave a variety of phenotypes. The $pd1$, $pd3$ and $pd4$ alleles gave class III incompatible transformants (--- type), characterized by severe growth inhibition with an average growth rate of 0.5 – 0.7 cm per day (Figure 3-17), no conidiation, and the occurrence of dead hyphal compartments after 24 hr of growth (Figure 3-19). Unlike standard $het-c$ class III transformants (such as introduction of $het-c^{OR}$ into a $het-c^{PA}$ strain), colonies of these class III transformants did not display a typical curly-edged morphology (Figure 3-18). The $pd2$ allele gave class II transformants (intermediate incompatibility), which displayed an average growth rate of 1.5 cm / day (Figure 3-17), little conidiation, and the occurrence of dead hyphae compartments at 3 days of vegetative growth (Figure 3-19 and Figure 3-18).

As shown in Table 3-12, transformants containing alleles $pd1$, $pd2$, $pd3$ exhibited different phenotypes from transformants containing standard $het-c^{OR}$, $het-c^{PA}$ or $het-c^{GR}$.
alleles. Allele $pd4$ showed specificity similar to $het-c^{GR}$. However, introduction of allele $pd4$ into a $het-c^{OR}$ strain gave transformants that exhibited a class III phenotype (severe incompatibility); introduction of the $het-c^{GR}$ into a $het-c^{OR}$ strain gave transformants that exhibited only a class II phenotype (intermediate incompatibility) (Table 3-12). The severity of the incompatible reaction may be due to amino acid differences in the specificity domain between $pd4$ and $het-c^{GR}$ (Figure 3-16).

Table 3-12. Summary of the phenotypes of the transformants of $het-c$ alleles $p26$, $pd1$, $pd2$, $pd3$ and $pd4$.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Phenotype of transformants in recipient strains of different $het-c$ specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$het-c^{PA}$</td>
</tr>
<tr>
<td>$het-c^{PA}$</td>
<td>++</td>
</tr>
<tr>
<td>$het-c^{OR}$</td>
<td>--</td>
</tr>
<tr>
<td>$het-c^{GR}$</td>
<td>-</td>
</tr>
<tr>
<td>$p26$</td>
<td>-</td>
</tr>
<tr>
<td>$pd1$</td>
<td>-</td>
</tr>
<tr>
<td>$pd2$</td>
<td>-</td>
</tr>
<tr>
<td>$pd3$</td>
<td>-</td>
</tr>
<tr>
<td>$pd4$</td>
<td>-</td>
</tr>
</tbody>
</table>

++, -, and -- represent the same phenotype as interpreted in Table 3-5.
Figure 3-17. Average growth rate comparison of \(pd1, pd2, pd3\) and \(pd4\) chimeric constructs transformed into PA (\(het-c^{PA}\)), OR (\(het-c^{OR}\)) and GR (\(het-c^{GR}\)) recipient strains. Bar indicates standard deviation. All the experiments were done in triplicates. PA, \(het-c^{PA}\). OR, \(het-c^{OR}\). GR, \(het-c^{GR}\).

<table>
<thead>
<tr>
<th>Alleles</th>
<th>PA</th>
<th>pd1</th>
<th>pd2</th>
<th>pd3</th>
<th>pd4</th>
<th>GR</th>
<th>OR</th>
</tr>
</thead>
</table>

Recipient strains
- into PA
- into OR
- into GR

Figure 3-18. Phenotypes of vegetative growth of transformants containing \(pd1, pd2, pd3\) and \(pd4\) constructs. Genotypes of the recipient strains are shown in Table 3-1. OR, \(het-c^{OR}\). PA, \(het-c^{PA}\). GR, \(het-c^{GR}\).
### Recipient strains

<table>
<thead>
<tr>
<th>Allele</th>
<th>( het-c^{PA} )</th>
<th>( het-c^{OR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( pd1 )</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>( pd2 )</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>( pd3 )</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>( pd4 )</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Figure 3-19.** Hyphal morphology and occurrence of dead hyphal compartments when \( pd1 \), \( pd2 \), \( pd3 \) and \( pd4 \) constructs were introduced into \( het-c^{PA} \) (C9-2 or Xa-3) and \( het-c^{OR} \) (C2-2-9 or C15-1) strains. Hyphae were examined within 1 day of growth. Dead hyphal compartments were observed by staining with 1% Evan’s blue for 5 min. Magnification 78X.
3.3.8 Generation of artificial alleles with both PA-specific and OR-specific insertions

To further investigate how the pattern of insertion affects het-c specificity, two artificial alleles pol and po2, with both OR-specific and PA-specific insertions, were constructed. Allele pol was generated by the two sets of primers red-pol.3 and pol.5-spl.3 using het-c\textsuperscript{OR} and het-c\textsuperscript{PA} as templates, respectively, for the first round of PCR. The same strategy was used to generate allele po2 except that primer sets were red-po2.3 and po2.5-spl.3 for the first round PCR. The pol and po2 alleles have the same insertion in the specificity region. The pol allele has the conserved NNG sequences from het-c\textsuperscript{OR} allele while po2 has the ENR sequences from the het-c\textsuperscript{PA} allele (Figure 3-20). The oligonucleotides pol.5 and po2.5 have a single nucleotide synonymous substitution to eliminate the Stu I restriction site in the PA-specific insertion region. This alteration allowed the identification of these two alleles in chimeric constructs. The DNA sequences of both alleles were confirmed (appendix 7.2).

The introduction of both pol and po2 alleles into het-c\textsuperscript{PA}, het-c\textsuperscript{OR}, and het-c\textsuperscript{GR} strains gave transformants displaying incompatible phenotypes (Table 3-14, Figure 3-21). Both het-c\textsuperscript{PA} and het-c\textsuperscript{GR} strains containing either pol or po2 constructs displayed class III incompatible phenotypes, described as severely inhibited growth (linear growth rate < 0.5 cm / day, Figure 3-22) and dead hyphal compartments after 24 hr of growth (Figure 3-23). The het-c\textsuperscript{OR} strain that contained either pol or po2 constructs showed class II incompatible phenotype.
Figure 3-20. Amino acid sequence comparison of PO1 and PO2 with HET-c\textsuperscript{PA}, HET-c\textsuperscript{OR} and HET-c\textsuperscript{GR}.

It is possible that the incompatible phenotype resulted from self-incompatibility of
pol or po2. However, the introduction of pol and po2 individually into a het-c null strain
X22-2 (het-\textsuperscript{6OR} het-c\textsuperscript{null} thr-2 a) gave only wild-type transformants (data not shown). These
wild type transformants displayed het-c mediated vegetative incompatibility in heterokaryon
tests with tester strains C3cr-29 (ro-7 un-24 het-\textsuperscript{6OR} het-c\textsuperscript{PA} pyr-4; inl a) and FGSC4571 (un-
24 het-\textsuperscript{6OR} het-c\textsuperscript{OR}; ad-3A; nic-2; cyh-1 A\textsuperscript{m54}) (Table 3-2) (data not shown). These data
indicate that pol and po2 confer novel het-c specificities that are different from het-c\textsuperscript{PA}, het-
c\textsuperscript{OR}, and het-c\textsuperscript{GR}.

To test whether pol and po2 confer identical or different het-c specificities, chimeric
constructs of pol and po2 with different selectable markers (either hyg\textsuperscript{R} or pan-2\textsuperscript{+}) were co-
transformed into a het-c null strain CJ44 (het-\textsuperscript{6OR} het-c\textsuperscript{null}; pan-2; arg-5 A). Only class I
(compatible) transformants were obtained, displaying an average growth rate of 4.7 cm per
day (see Table 3-15 in section 3.3.11). These data indicate that pol and po2 confer the same
het-c specificity, and that the specificity of pol and po2 is not affected by the amino acid
residue differences in the NNK or ENR block (Figure 3-20).
Table 3-13. Summary of transformant phenotypes containing *pol* and *po2* chimeric constructs.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Phenotype of transformants in recipient strain of different <em>het-c</em> specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>het-c</em>!&lt;sup&gt;GR&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>het-c</em>!&lt;sup&gt;GR&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td><em>het-c</em>!&lt;sup&gt;PA&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td><em>het-c</em>!&lt;sup&gt;OR&lt;/sup&gt;</td>
<td>++</td>
</tr>
<tr>
<td><em>pol</em></td>
<td>-</td>
</tr>
<tr>
<td><em>po2</em></td>
<td>-</td>
</tr>
</tbody>
</table>

Classification of -, -- and ++ was described in Table 3-5.
Figure 3-21. Phenotypes of transformants of mutant alleles pol and po2 chimeric constructs. Genotypes of the recipient strains are shown in Table 3-1. OR, het-c^{OR}. PA, het-c^{PA}. GR, het-c^{GR}.

Figure 3-22. Average growth rate of transformants containing alleles pol and po2. Bar indicates standard deviation. Recipient strains: PA, het-c^{PA} (C9-2 or Xa-3); OR, het-c^{OR} (C2-2-9 or C15-1); GR, het-c^{GR} (FGSC2193). All experiments were done in triplicates. PA, het-c^{PA}. OR, het-c^{OR}. GR, het-c^{GR}.
**Figure 3-23.** Hyphal compartmentation and death in OR \((het-c^{OR})\) and PA \((het-c^{PA})\) transformants containing either \(pol\) and \(po2\) constructs. The presence of dead hyphal compartments was assayed after 1 day of growth by staining with 1% Evan’s blue for 5 min. Magnification 82X.
3.3.9 Alleles with chimeric combinations of amino acid sequence and patterns of insertion/deletion.

Two polymorphic regions flanking the insertion/deletion motif are highly variable between het-c allele types (regions I and II in Figure 3-1). Genetic analysis (Saupe and Glass, 1997) and studies on naturally occurring het-c alleles indicated that the amino acid composition of polymorphic region I, MEERRGGH/IHE(Q)KET(D)RS, does not affect het-c allelic specificity. To further investigate whether the amino acid composition of polymorphic region II, TDTRVTL/TATK(Q)L(T)L, affects het-c allelic specificity, alleles with artificial combinations of polymorphic regions and patterns of insertion/deletion were constructed (po3, po4, odl and gol, Figure 3-24). One type, po3 and po4, contains mosaic combinations between het-c^{PA} and het-c^{OR}. The allele po3 has a PA-specific insertion with an OR-specific polymorphic region II, TDTRVTL, while po4 has an OR-specific insertion with a PA-specific polymorphic region II, TATKLKL. The second type of alleles (odl and gol) contains mosaic combinations between het-c^{GR} and het-c^{OR}. Allele odl was generated from het-c^{OR} with a deletion of the OR-specific insertion sequence, RNDTR; while allele gol was generated from het-c^{GR} with an addition of the OR-specific insertion sequence, RNDTR (Figure 3-24). The odl and gol chimeric constructs were identified by the unique Apa I site within the OR-specific insertion sequence.
Figure 3-24. Sequence comparison of the variable domain of PO3, PO4, OD1 and GO1. Bold region indicates polymorphic block II of HET-c<sup>OR</sup> and HET-c<sup>PA</sup>.

Transformation assays showed that po3 and po4 chimeric constructs have the same specificity as wild type het-c<sup>PA</sup> and het-c<sup>OR</sup> constructs, respectively (Table 3-14). Introduction of the po3 chimeric construct into the het-c<sup>OR</sup> strain gave class II type of transformants that resemble the phenotype of a het-c<sup>OR</sup> strain containing a het-c<sup>PA</sup> allele (Table 3-14). Only class I transformants were obtained when the po3 chimeric construct was introduced into either het-c<sup>PA</sup> or het-c<sup>GR</sup> strains (data not shown). The introduction of the po4 chimeric construct into het-c<sup>PA</sup> and het-c<sup>GR</sup> strains gave class III type of transformants. The phenotype of these transformants was similar to that of the introduction of the het-c<sup>OR</sup> construct into het-c<sup>PA</sup> and het-c<sup>GR</sup> strains. The introduction of either po4 chimeric constructs into het-c<sup>OR</sup> strain only gave class I transformants (data not shown).

The introduction of odl and gol chimeric constructs gave transformants that had phenotypes resembling transformants containing het-c<sup>GR</sup> and het-c<sup>OR</sup> constructs, respectively (Table 3-14). The introduction of the odl chimeric construct into het-c<sup>OR</sup> and het-c<sup>PA</sup> strains both gave class II type of transformants (intermediate incompatibility) displaying an average growth rate of 1.4 cm per day; only class I (compatible) transformants were obtained when the odl chimeric construct was introduced into het-c<sup>GR</sup> strain. Introductions of gol chimeric constructs into het-c<sup>PA</sup> and het-c<sup>GR</sup> strains both gave class III type of transformants, displaying a severe incompatibility phenotype and an average growth rate of 0.3 cm per day;
while introduction of *gol* chimeric construct into *het-c*\textsuperscript{OR} strain only gave class I (compatible) transformants displaying an average growth rate of 4.9 cm per day.

**Table 3-14.** Summary of the phenotype of transformants containing *po3*, *po4*, *odl* and *gol* chimeric constructs. Classification of ++, - and -- was described in Table 3-5.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Phenotype of transformants in recipient strains of different <em>het-c</em> specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>het-c</em>\textsuperscript{OR}</td>
</tr>
<tr>
<td><em>het-c</em>\textsuperscript{OR}</td>
<td>++</td>
</tr>
<tr>
<td><em>het-c</em>\textsuperscript{PA}</td>
<td>-</td>
</tr>
<tr>
<td><em>het-c</em>\textsuperscript{GR}</td>
<td>-</td>
</tr>
<tr>
<td><em>po3</em></td>
<td>-</td>
</tr>
<tr>
<td><em>po4</em></td>
<td>++</td>
</tr>
<tr>
<td><em>odl</em></td>
<td>-</td>
</tr>
<tr>
<td><em>gol</em></td>
<td>++</td>
</tr>
</tbody>
</table>

**3.3.10 Summary of the *het-c* specificity of artificial alleles**

Artificial chimeric *het-c* allele construction and transformation assays indicated that novel *het-c* specificities could be generated by altering the pattern of insertion (i.e., *pol/ppo2*) or the size of the PA-specific insertion (i.e. *pd1*, *pd2* and *pd3*) in the variable domain and that *het-c* allelic specificity was not affected by the exchange of the polymorphic regions between alleles (i.e. *po3*, *po4*, *odl* and *gol*). Alleles *pol/ppo2*, *pd1*, *pd2* and *pd3* displayed incompatibility with strains containing wild-type *het-c* alleles, *het-c*\textsuperscript{OR}, *het-c*\textsuperscript{PA} and *het-c*\textsuperscript{GR}.
although the severity of incompatibility reactions differ. Whether pol/po2, pd1, pd2 and pd3 are mutually incompatible will be described in section 3.3.12.

3.3.11 Deletion disrupting het-c activity

An allele del3 was generated by PCR from het-cGR that contained a deletion of the three amino acids NNG in the variable domain (Figure 3-25). The Stu I − Sal I fragment of the PCR product was cloned into het-cPA (StuI−SalI) vector cassette and del3 was identified by the absence of an Xho I site (data not shown). The DNA sequence of the del3 construct was verified by sequence analysis (see Appendix 7.2). Only class I transformants were obtained when the del3 chimeric construct was introduced into het-cOR, het-cPA or het-cGR recipient strains (Figure 3-26); hyphal compartmentation and death were not observed in any of the transformants (data not shown). These transformants had comparable growth rates with those of wild type het-cOR, het-cPA and het-cGR transformants, respectively (Figure 3-27). These data suggest that the three amino acids NNG are required for HET-c function, protein folding or stability. A change of the sequence NNG to ENR, however, does not affect het-c specificity as shown in the construction of alleles pol and po2 (section 3.3.8).

Figure 3-25. Sequences comparison of DEL3, HET-cGR, HET-cOR and HET-cPA. DEL3 was generated from HET-cGR by deletion of NNG.
**Figure 3-26.** Phenotypes of transformants containing del3 construct. Genotypes of the recipient strains are shown in Table 3-1. OR, het-c^{OR}. PA, het-c^{PA}. GR, het-c^{GR}.

**Figure 3-27.** Average growth rate (cm/day) of transformants containing del3 construct. Bar indicates standard deviation. Recipient strains: PA (het-c^{PA}), C9-2 or Xa-3; OR (het-c^{OR}), C2-2-9 or C15-1; GR (het-c^{GR}), FGSC2193. Alleles: PA, het-c^{PA}, OR, het-c^{OR}, GR, het-c^{GR}, del, del3. All experiments were done in triplicates.
3.3.12 Co-transformation of alleles with novel specificities

The artificial alleles, pdl, pd2, pd3 and pol/po2 (pol and po2 confer the same specificity, see section 3.3.8), confer a different specificity from $het-c^PA$, $het-c^{OR}$ and $het-c^{GR}$ alleles (Table 3-12 and 3-13). Do alleles pdl, pd2, pd3 and pol/po2 confer the same specificity or display mutual incompatibility? To address this question, pair-wise combinations of these artificial alleles were co-transformed into a $het-c$ null strain CJ44 ($het-c^{6R}$ $het-c^N$; pan-2; arg-5 A) to assay incompatibility. Wild type $het-c^PA$, $het-c^{OR}$ or $het-c^{GR}$ constructs and the vectors pCB1004 and pOKE103 were used as controls.

To determine if the vectors used in $het-c$ chimeric construction affected the phenotypes of co-transformants, control experiments with vectors pCB1004 and pOKE103 were performed. Transformants containing vectors pCB1004 and pOKE103 (by co-selection of hyg$^R$ and PAN requirement) displayed class I phenotypes (Table 3-5), with vigorous growth, conidiating and an average growth rate of 4.6 cm / day and no occurrence of dead hyphal compartments (Table 3-15).

To examine if the recipient strain CJ44 is an authentic $het-c$ null strain, constructs of $het-c$ alleles with defined specificity (i. e. $het-c^{OR}$, $het-c^PA$ or $het-c^{GR}$) were co-transformed into strain CJ44 by co-selection of hyg$^R$ and PAN requirement. Transformants containing constructs of the same $het-c$ alleles (i. e. $het-c^{OR}$/het-c$^{OR}$, or $het-c^PA$/het-c$^{PA}$) displayed a class I phenotype (compatible; Table 3-15). Transformants containing constructs of different $het-c$ alleles, i. e. $het-c^PA$/het-c$^{OR}$ or $het-c^{GR}$/het-c$^{OR}$, exhibited class III phenotypes (severely incompatible, Table 3-15), with curly, aconidial morphology, average growth rate of 0.3 cm per day and dead hyphal compartments within one day of growth. These results were consistent with $het-c^{OR}$/het-c$^{PA}$ and $het-c^{OR}$/het-c$^{GR}$ partial duplication results (Saupe and
Glass, 1997). These experiments confirmed that the recipient strain *CJ44* is a *het-c* null strain and that the co-transformation assay is working.

To test if alleles *pdl*, *pd2*, *pd3* and *po1/po2* are self-incompatible, two different constructs (in *pCB1004* vector and in *pOKE103* vector) of the same allele were cotransformed into the *CJ44* recipient strain to assay incompatibility. All of the "self" co-transformants exhibited class I (compatible) phenotypes, with an average growth rate of 4.3 – 5.4 cm / day and no occurrence of dead hyphal compartments (Table 3-15).

As described in section 3.3.8, transformants containing constructs of alleles *pol* and *po2* showed class I phenotype (compatible phenotype), indicating that *pol* and *po2* conferred the same *het-c* specificity. Transformants containing constructs of pair-wise combinations of artificial *het-c* alleles, e.g., *pol/pdl*, *pol/pd2*, *pol/pd3*, *pd1/pd2*, *pd1/pd3*, or *pd2/pd3*, were co-selected for *hyg* R and PAN requirements. All the transformants displayed a class III phenotype, showing severely inhibited growth (with an average growth rate of 0.3 – 0.6 cm per day), aconidial morphology and occurrence of dead hyphal compartments within one day of vegetative growth (Table 3-15). These results indicated that *po1/po2*, *pdl*, *pd2* and *pd3* each confers a different *het-c* specificity. Thus, a number of new *het-c* specificities were generated by altering the pattern of the insertion or the size of the PA-specific insertion in the specificity domain.

In the sequence comparison of all the *het-c* alleles, the artificial allele *pdl* shares the same pattern and size of insertion in the specificity motif as the natural allele *p26*, although the amino acid compositions in the specificity region are variable (Figure 3-28). To examine if *pdl* and *p26* confer the same specificity, chimeric constructs of *pdl* and *p26* were co-
transformed into the CJ44 strain. Only compatible transformants were obtained, suggesting that pd1 and p26 confer the same het-c specificity.

The polypeptides encoded by the alleles pd3 and het-cOR have the same size of the variable domain but have different amino acid sequences (Figure 3-29). The PD3 has five amino acid insertion, QFRRK, in the PA-specific motif, while the HET-cOR has a five amino acid insertion, RNDTR, in the OR-specific motif. Transformants containing both pd3 and het-cOR constructs displayed a class III (severe incompatibility) phenotype, which is consistent with the results observed when the pd3 construct was introduced into a het-cOR strain (Table 3-13).

P26      ALIDIHQKEDRSKSRIPPHVGTRTKLKRENQMFVPAFYERDYYWPLVT  
PD1      VLIDMEERRGGH-SPVFPHVTATKLKLENQFRVRPGEGKYAWPLVT  
          ***    **    ***    ****    **    *    *    *    ***

Figure 3-28. Sequence comparison of P26 and PD1. Asterisks indicate identical sites.

OR       VLIDMEERRGGHSPVFPHGTATKLKLENQFRVRPGEGKYAWPLVT  
PD3      VLIDMEERRGGHSPVFPHGTATKLKLENQFRVRPGEGKYAWPLVT  
          ******************    *    *    *    *    ****

Figure 3-29. Sequence comparison of PD3 and OR. Asterisks indicate identical sites.
Table 3-15. Phenotypes of co-transformants of pol/po2, pd1, pd2 and pd3 chimeric constructs into strain CJ44.

<table>
<thead>
<tr>
<th>Allele pairs</th>
<th>incompatibility</th>
<th>Growth Rate (cm / day) (mean ± SD)</th>
<th>Hyphal compartmentation and death observed by Evan's blue staining</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCB + POKE</td>
<td>++</td>
<td>4.6 ± 0.20</td>
<td>Not significant</td>
</tr>
<tr>
<td>het-cGR + het-cGR</td>
<td>++</td>
<td>4.7 ± 0.20</td>
<td>Not significant</td>
</tr>
<tr>
<td>het-cPA + het-cPA</td>
<td>++</td>
<td>4.8 ± 0.20</td>
<td>Not significant</td>
</tr>
<tr>
<td>het-cGR + het-cPA</td>
<td>--</td>
<td>0.8 ± 0.10</td>
<td>Significant within 1 day of growth, more at 3 days of growth</td>
</tr>
<tr>
<td>het-cGR + het-cGr</td>
<td>--</td>
<td>0.4 ± 0.10</td>
<td>Significant within 1 day of growth, more at 3 days of growth</td>
</tr>
<tr>
<td>pd1 + pd1</td>
<td>++</td>
<td>4.3 ± 0.30</td>
<td>Not significant</td>
</tr>
<tr>
<td>pd2 + pd2</td>
<td>++</td>
<td>5.4 ± 0.60</td>
<td>Not significant</td>
</tr>
<tr>
<td>pd3 + pd3</td>
<td>++</td>
<td>4.4 ± 0.30</td>
<td>Not significant</td>
</tr>
<tr>
<td>p26 + pd1</td>
<td>++</td>
<td>4.5 ± 0.35</td>
<td>Not significant</td>
</tr>
<tr>
<td>p26 + pd2</td>
<td>++</td>
<td>4.6 ± 0.17</td>
<td>Not significant</td>
</tr>
<tr>
<td>p26 + p26</td>
<td>++</td>
<td>4.6 ± 0.17</td>
<td>Not significant</td>
</tr>
<tr>
<td>pd3 + het-cGR</td>
<td>--</td>
<td>0.6 ± 0.15</td>
<td>Significant within 1 day of growth, more at 3 days of growth</td>
</tr>
</tbody>
</table>

Not significant means hyphal compartmentation and death < 1% in a population. Significant hyphal compartmentation and death > 10% of a population. All the experiments were done in triplicates. Description of ++ and -- phenotype is in Table 3-5.
3.4 DISCUSSION

The difference in a variable region of 34 - 48 amino acids was shown to be sufficient to alter *het-c* specificity as previously shown by chimeric construction (Saupe and Glass, 1997). However, factors important for *het-c* specificity in the variable region were not determined. A peptide sequence comparison of the variable region showed two polymorphic blocks (I and II in Figure 3-1), two independent insertion or deletion motifs (OR-specific five amino acid insertion and PA-specific 14 - 16 amino acid insertion) and variation in amino acid sequences among alleles (Figure 3-1). Genetic data from partial diploid analysis indicated that the polymorphic block did not contribute to *het-c* allele specificity (Saupe and Glass, 1997). In this study, I further investigated the determinants for *het-c* allele specificity via chimeric allele construction and *in vitro* mutagenesis. The data will help to gain a general understanding of how *het-c* allele specificity is determined and how recognition may be mediated.

3.4.1 Variation in *het-c* mediated incompatibility phenotypes

Genetic and mutational analysis suggested that vegetative incompatibility reactions resulted from the co-expression of different *het-c* alleles. Theoretically, most of the cells should have identical nuclei in any single transformant and thus should display a nearly identical incompatibility phenotype. Although hyphal growth rate and conidiation were affected throughout the colony, only approximately 30% of the hyphal segments displayed compartmentation and death. This phenomenon was also observed in incompatible *het-c* partial diploids (Jacobson et al., 1998).
Several possible interpretations could account for these observations. First, co-expression of different het-c alleles may not be the only factor essential for inducing vegetative incompatibility reaction. Other modifier molecules may be required to facilitate the activity of the het-c products. If the expression or/and subcellular compartmentation of these modifier molecules varies among hyphal segments, the degree of the incompatibility reaction may also vary. The second scenario is related to the stability of the het-c products, which could be affected by the physiology of the hyphal compartment. Third, a threshold level of the different het-c products could be required for triggering an incompatibility reaction, and this level may vary among hyphal segments. Whether the expression of het-c is regulated during different vegetative growth phases is unknown due to the difficulty of detecting het-c expression by Northern analysis (this study and S. J. Saupe, personal communication).

In the class II transformants (intermediate incompatibility), hyphal compartmentation and death do not occur until after two to three days of vegetative growth, however, growth was significantly inhibited. This phenotype could be explained by a modifier in the recipient strains that may interact with the components involved in the incompatible reaction. For example, introduction of the het-c<sup>GR</sup> allele into a het-c<sup>PA</sup> strain gave class II (intermediate) incompatible transformants while introduction of the het-c<sup>PA</sup> allele into a het-c<sup>GR</sup> strain gave class I (compatible) transformants (this study and S. J. Saupe, personal communication). Since the het-c<sup>GR</sup> strain used in this study is not isogenic at other loci (L. Glass, personal communication), it is possible that differences in specific modifier in the het-c<sup>GR</sup> and het-c<sup>PA</sup> recipient strains may cause phenotypic differences in the transformants.
The effects of genetic background or modifier in the recipient strains on the phenotypes of transformants were also shown by comparisons of individual transformants and co-transformants of alleles pol, pdl, pd2, or pd3 with het-c^{PA}, het-c^{OR} or het-c^{GR} (data not shown). For example, co-transformants of het-c^{OR} + pol, het-c^{OR} + pd2, het-c^{GR} + pd2, and het-c^{GR} + pd3, displayed class III (severe incompatibility) phenotypes while individual transformants of either pol and pd2 into a het-c^{OR} strain or pd2 and pd3 into a het-c^{GR} strain gave only class II (intermediate incompatibility) phenotypes as shown in Table 3-12 and 3-13. One explanation for the display of different incompatible phenotypes could be due to the different genetic background between the two transformation systems, in particular the background interference (or modifications) of the recipient het-c^{OR} and het-c^{GR} strains. However, it is possible that the more severe incompatible phenotype in het-c co-transformants resulted from the high copy numbers and expression of het-c constructs.

3.4.2 het-c allele specificity was not affected primarily by amino acid substitutions in the variable region

I first addressed the question whether the amino acid sequence composition in the variable region was the critical determinant for het-c allele specificity using the 24 naturally occurring het-c alleles. Twenty-three of the 24 naturally occurring alleles fall into one of the three categories of het-c allelic types as designated by the pattern of insertions or deletions first observed in N. crassa (Saupe and Glass, 1997). Within each allelic group, alleles differ in their amino acid sequence at multiple sites. An exception to the 24 naturally occurring alleles, Ndi5923 (p26) has a different insertion size but the same insertion pattern as PA-type alleles.
Specificity of the 24 naturally occurring alleles was assayed by chimeric allele construction and transformation in *N. crassa* strains of different *het-c* constitution. The results indicated that the alleles with an OR-specific insertion pattern (insertion of five amino acid residues RNDTR from position 54 to 59) or GR-specific deletion pattern (with neither PA-specific insertion or OR-specific insertion) conferred the same specificity as wild type *het-c*<sup>OR</sup> (*Nc2489*) or *het-c*<sup>GR</sup> (*Nc1945*) allele, respectively. Thus, the amino acid diversity that occurs between OR-type and GR-type of alleles does not alter *het-c* specificity.

The PA-type alleles have the same overall insertion pattern (the numbers of amino acid residues vary from 14 - 16), but the amino acid sequences in the insertion motif vary among alleles. Of the seven PA-type alleles, three alleles (*Np7221*, *Gsp8243* and *Ss2740*) have 16 amino acid residues in the insertion motif, two alleles (*Nc1130* and *Nt6583*) have 15 amino acid residues in the insertion motif, and two (*Nc4832* and *Nc2316*) have only 14 amino acid residues in the insertion motif. However, all seven alleles conferred the same specificity as the wild type *het-c*<sup>PA</sup> allele (*Nc1130*, Saupe and Glass, 1997).

The *Ndi5923* (*p26*) allele has only 11 amino acid residues in the PA-specific insertion motif. The *p26* allele confers a *het-c* specificity that is different from the previously known three types of *het-c* alleles, *het-c*<sup>OR</sup>, *het-c*<sup>PA</sup> and *het-c*<sup>GR</sup>. The novel *het-c* specificity of *p26* could be due to either the different insertion size or the amino acid sequence variation or the combined effect of both in the insertion motif of P26. However, P26m, which has additional five amino acid residues to the specific insertion motif of P26, confers the same specificity as that of other *het-c*<sup>PA</sup> type of alleles. These data suggest that the specificity of *het-c* is affected by the size of the insertion rather than by the amino acid composition in the PA-insertion motif. This conclusion is further supported by a comparison of the allele pairs, *p26* and *pd1*. 
that confer the same het-c specificity. Both P26 and PD1 have the same pattern and size of insertion, but the amino acid compositions in the insertion motifs are different. These results indicate that the novel specificity of p26 resulted from the reduction in the insertion size rather than amino acid sequences in the insertion motif, supporting the observation that amino acid sequence composition is not the primary determinant of het-c allele specificity.

However, since information is not available regarding how conserved amino acid residues in the variable domain affect HET-c function, and because the specificity of alleles with random compositions of the amino acid sequences in the variable domain was not tested, it is not possible to conclude that amino acid composition in the variable domain is not entirely irrelevant for HET-c function.

3.4.3 The pattern and size of insertion or deletion in the variable region are the primary determinants for het-c allelic specificity

Studies on the specificity of allele p26 suggested that it is possible to generate a set of novel het-c specificity types by reducing the size of the PA-specific insertion. In this study, novel het-c specificities were generated either by alterations in the pattern of insertion (allele pol/po2) or by alterations in the size of the PA-insertion motif (alleles pd1, pd2 and pd3). Sequence comparison suggested that the new specificity could result from a particular pattern and size of the insertions/deletions in the variable domain or from alterations in the spacing (the numbers of amino acid residues) in the variable domain. Interestingly, PD3 and HET-cOR have the same number of amino acid residues in the variable domain (Figure 3-29) but allele pd3 confers a different specificity from het-cOR, suggesting that new specificities could be generated by amino acid substitutions in the variable domain. It is interesting to note that
alterations in the size of the PA-insertion motif do not always result in novel het-c specificity. For instance, Nc4832 is one amino acid shorter than Nc1130 and two amino acid residues shorter than Gsp8243 in the PA-specific insertion; however, alleles Nc4832, Nc1130 and Gsp8243 all confer the same specificity. In contrast, PD2 is two amino acid residues shorter than PD1; alleles pd1 and pd2 confer different specificities. How the specificities were generated and how the size of PA-specific insertion may affect het-c allelic specificity will be discussed in Section 3.4.5.

3.4.4 Predicted protein structural characteristics of the hypervariable region

The specificity domain of het-c is rather hydrophobic and contains short stretches of amino acids, such as sequence FPHVG that are conserved among all alleles. The secondary structures of HET-c^{OR}, HET-c^{PA} and HET-c^{GR} in this domain were predicted hypothetically by the Gibrat (Gibrat et al., 1987), Levin, (Levin et al., 1986), DPM (Deleage and Roux, 1987) and SOPMA (Geourjon and Deleage, 1994; 1995) prediction programs (http://www.ibcp.fr) and are shown in Figure 3-30. These programs have been shown to have accuracy rates of approximately 68 – 70 % (Gilbrat et al., 1987; Levin et al., 1986; Deleage and Roux, 1987; Geourjon and Deleage, 1994; 1995). All these programs are based on algorithms of prediction to assign secondary structures, using the homology between fragments of submitted sequence and sequences in the database of protein conformations.

Secondary structures of the HET-c specificity domain predicted by the different methods agree with each other although the prediction methods use very different algorithms. In general, the predicted consensus structure of this domain for all three HET-c peptides is characterized by a Helix - Coil - Helix - Coil- β-Sheet – Coil - β-Sheet motif (Figure 3-30),
except that an extra helix structure is predicted in HET-c\textsuperscript{PA} near the PA-specific insertion. The highly conserved region upstream of the hypervariable domain (helix I and II in Figure 3-30) in all three HET-c peptides is predicted to form a helical structure by all the methods. The conserved region downstream of the hypervariable domains (β-strand II) in all three HET-c peptides is predicted to form a β-sheet structure. The two specific insertion motifs, PA-specific and OR-specific, are both predicted to form a coil structure between two β-strands. In the case of HET-c\textsuperscript{GR}, the conserved four residues NNGK are predict to form a coil structure between two β-strands, which is similar to that of the OR-specific or PA-specific insertion motif.
Figure 3-30. Secondary structure prediction for HET-c specificity region. Helix - H, Beta sheet - E, Coil – C, Turn – T. Bold sequences indicate PA, OR, and GR-specific sequences. Underlined region is the variable region (v). Asterisks indicate hydrophobic residues.
3.4.5 Model for generation of *het-c* allele specificity

Most of the residues in the helices and the β-strands have clustered hydrophobic side chains in all three HET-c polypeptides (Figure 3-30, indicated by asterisks). Hydrophobic side chains are usually in the interior of protein molecules. It is possible that in the specificity region of HET-c, the two β-strands may pack tightly together in an antiparallel configuration through side chain hydrophobic interactions to form a stable framework of protein conformation (Figure 3-31). The allele-specific variable domain (the coil structure) may form a specific protruding loop (or a groove) to cover the surface of the compact structure. The two anti-parallel β-strands and the allele-specific loop (or groove) may thus form a specific conformation that mediates *het-c* allele specificity (Figure 3-31). The coil loop (or groove) structure may vary between different *het-c* allelic types but is similar within each allelic group. The variable loop domain may be involved in protein-protein interactions (between HET-c proteins or with other proteins) or may regulate the stability of protein complex formation. Similar structures have been well demonstrated by crystalline studies with immunoglobulin molecules where the antigen interface consists of hypervariable loops formed by two tightly antiparallel β-strands known as the complementarity determining region (CDR) (Davis et al., 1990; Creighton, 1993). Such a structure has also been shown in the T-cell receptor recognition surface of MHC class I molecules (Jones et al., 1998; Zhang et al., 1998).

The HET-c\textsuperscript{OR} - type peptides are predicted to have the same secondary structure in the specificity region (data not shown), and thus the same conformations in the variable domain. All the HET-c\textsuperscript{GR} - type proteins are predicted to have the same conformations in the specificity domain as well (data not shown). In the case of HET-c\textsuperscript{PA} - types, all the
polypeptides are predicted to form similar secondary structure, although the amino acid sequences in the PA-specific insertion motif are highly variable (data not shown). It is possible that all the het-c$_{PA}$ - type proteins would have similar conformations in the variable domain as well. Therefore, explicitly, within each het-c allelic group, the proteins are predicted to have the same or similar conformations in the variable domain.

The hypothesis also explains the generation of new het-c allele specificities. Alterations in the pattern of insertion (alleles pol and po2) or size of the PA-insertion motif (alleles pd1, pd2 and pd3) would result in the formation of a different loop domain (the V region/domain in Figure 3-30 and 3-31) between the two β-strands and thus new conformations in the variable domain that confer a new allele specificity. A deletion of the NNG from HET-c$_{GR}$ (DEL3) results in the loss of the allele-specific coil region (underlined in Figure 3-30), which would explain the elimination of its allele specificity.

An artificial allele pd3 results in a polypeptide that has the same length of amino acid residues in the variable region as HET-c$_{OR}$. However, allele pd3 confers a different specificity from het-c$_{OR}$ as shown by transformation assay (section 3.3.7). It could be explained by the differences in the predicted secondary structures of PD3 and HET-c$_{OR}$ (Figure 3-32). According to the predicted secondary structure, the conformation of PD3 in the variable domain would result in two shorter antiparallel β-strands and a short protruding coil structure of four amino acid residues, which differs from that of HET-c$_{OR}$.

Differences in the size (by one or two amino acids) of the PA-specific insertion in naturally occurring het-c$_{PA}$ - type alleles did not result in different het-c specificities. One could explain this by postulating that the transformation assay is not discriminating enough to differentiate these different het-c specificities. Alternatively, if the variable loop domain is
involved in protein-protein interactions or regulating the stability of protein complex formation, it could be explained by the strength of the interactions mediated via the variable loop domain. It is observed that the het-c\textsuperscript{PA} - type alleles have either four or five hydrophobic residues in the variable loop (corresponding to the underlined region in Figure 3-30). If hydrophobic interaction is a force to stabilize protein complex formation, one would predict that all HET-c\textsuperscript{PA} - type proteins would mediate similar dynamics in protein-protein interactions.

Two hydrophobic sites, residues P/G and G, are conserved in all alleles (except Ndi5923). These two sites could be important for protein-protein hydrophobic interactions. This speculation could account for the different specificities of alleles pd1, pd2 and pd3. It is observed that PD2 has a deletion of the conserved residue, G, compared with PD1, and that PD3 has a deletion of the conserved hydrophobic residue, P, compared to PD2.

The model discussed here is hypothetical and is based on computer prediction and chimeric allele analysis. No structural data are available to support these hypotheses of how het-c specificity is generated and that the variable motifs are involved in protein-protein interactions. However, these hypotheses can be further tested experimentally.
Figure 3-31. Proposed protein conformations in the variable domain for determining HET-c specificity. The two anti-parallel β-strands (see Figure 3-30) may pack tightly together by hydrophobic interactions. The protruding loop structure on the surface is composed of the deletion/insertion region of HET-c polypeptide (see Figure 3-30). See text for details.
<table>
<thead>
<tr>
<th>sp-PD3</th>
<th>EALRCQLQALHTLDFFPASHNYCEVLIDMEERRGGSVPFPHVGTATKLKLERNRFPFVYAYPLVTGTFFG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>** **** * * *** * **** * * **** * * * * **** * * * * **** * * * * ****</td>
</tr>
<tr>
<td>Gibrat</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>Levin</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>DPM</td>
<td>CCHHHHHHHHHHHHHHHHHHCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHH</td>
</tr>
<tr>
<td>SOPMA</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>Consensus</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sp-OR</th>
<th>EALRCQLQALHTLDFFPASHNYCEVLIDMEERRGGSVPFPHVGTDTTRITLRNDTRNNGKSVWPLVTGTFFG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>** **** * * *** * **** * * **** * * * * **** * * * * **** * * * * ****</td>
</tr>
<tr>
<td>Gibrat</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>Levin</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>DPM</td>
<td>CCHHHHHHHHHHHHHHHHHHCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHHHCCHH</td>
</tr>
<tr>
<td>SOPMA</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
<tr>
<td>Consensus</td>
<td>HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>helix I</th>
<th>helix II</th>
<th>helix III</th>
<th>β-sheet I</th>
<th>V</th>
<th>β-sheet II</th>
</tr>
</thead>
</table>

**Figure 3-32.** Comparison of predicted secondary structure PD3 and HET-cOR in the specificity region. Helix - H, Beta strand - E, Coil - C, Turn - T. Asterisks indicate hydrophobic residues.
3.4.6 The use of artificial het-c alleles to study protein interactions

A number of artificial chimeric het-c alleles were constructed in this study for identifying the critical determinants for allele specificity. Whether the hypervariable region is involved in protein-protein interactions was not addressed in this study. Biochemical approaches using co-immunoprecipitation and genetical approaches such as yeast two-hybrid assays are required. However, these artificial and chimeric het-c alleles constructed in this study will be very useful for studying HET-c interactions via yeast two-hybrid system and in structural studies.
Chapter 4 Heteromeric complex formation of HET-c proteins

4.1 INTRODUCTION

Morphological features (Jacobson et al., 1998) and evidence of nuclear DNA degradation (Marek et al., 1999) suggested that the cellular responses of het-c mediated vegetative incompatibility have features in common with programmed cell death (PCD) in animal and plant cells. However, the genetic and biochemical mechanisms mediating the vegetative incompatibility response remain largely unknown.

Mechanisms mediating non-self recognition have been investigated in many biological systems, such as the mating systems in S. cerevisiae (Ho et al., 1994), S. commune (Specht et al., 1992) and U. maydis (Kamper et al., 1995), and the transcriptional activator Fos-Jun oncoproteins (O'Shea et al., 1992). Protein heterodimerization was commonly found as an important mechanism for mediating recognition in these systems. In P. anserina, it was also proposed that heteromeric complex formation of different het gene products is a mechanism triggering vegetative incompatibility response, although this hypothesis is based on genetic and mutational analyses (Begueret et al., 1994).

In N. crassa, a strain carrying a mutated het-c<sup>OR</sup> allele displayed dual compatibility with both het-c<sup>OR</sup> and het-c<sup>PA</sup> strains; introduction of a mutated het-c<sup>OR</sup> allele into a het-c<sup>PA</sup> strain only gave compatible transformants (Saupe et al., 1996b). These studies indicate that het-c mediated vegetative incompatibility reaction requires the co-expression of different het-c alleles. Different mechanisms could be proposed on how het-c mediates non-self
recognition during vegetative incompatibility. One hypothesis is that non-self recognition during vegetative incompatibility is mediated via HET-c heteromeric complex formation. Under this hypothesis, formation of a HET-c$_{OR}$ and HET-c$_{PA}$ heteromeric complex is associated with the incompatible phenotype of strains containing het-c$_{OR}$ and het-c$_{PA}$ alleles, regardless of whether homomeric complexes between HET-c protein themselves are formed or not. Alternatively, one could hypothesize that non-self recognition during vegetative incompatibility is not mediated via heteromeric complex formation but is a consequence of sequential regulatory events. In this case, one het-c product (e. g., HET-c$_{OR}$) may regulate the function or expression of a different het-c product (e. g., HET-c$_{PA}$) or other cellular factors, which may result in a dominant negative effect on vegetative growth; however, vegetative incompatibility in this case is not associated with HET-c$_{OR}$ and HET-c$_{PA}$ heteromeric complex formation.

To distinguish between these two possibilities, I performed immunoprecipitation assays to examine if HET-c$_{OR}$ and HET-c$_{PA}$ form a heteromeric complex. Whether HET-c$_{OR}$ forms a homomeric complex during vegetative growth was also examined. Because of the high similarity of putative het-c allele products (Saupe and Glass, 1997), two molecular probes were used in this study to detect different HET-c proteins: the green fluorescent protein (GFP) and the nine amino acid hemagglutinin (HA) tag. The GFP gene was fused with het-c$_{OR}$, and the HA gene was fused with both het-c$_{OR}$ and het-c$_{PA}$.

GFP and HA have been widely used as reporters in heterologous expression systems via gene fusion. The green fluorescent protein (GFP) is a spontaneously fluorescent 27 kDa protein (238 amino acids) originally isolated from the jellyfish Aequorea victoria. It has elicited much interest as a reporter in heterologous systems for in vivo visualization of protein
subcellular localization in a wide variety of cell types and organisms, such as *E. coli* (Cody et al., 1993), *Caenorhabditis elegans* (Chalfie et al., 1994), *Drosophila melanogaster* (Wang and Hazelrigg, 1994) and the filamentous fungus, *A. nidulans* (Sievers and Fischer, 1997). The amino acid sequence (YPYDVPDYA) of hemagglutinin, a protein of the influenza virus, has been successfully used to create epitope-tagged proteins for subcellular localization in *D. melanogaster* (Surdej and Jacobs-Lorena, 1994) and *A. nidulans* (Sievers and Fisher, 1997).

To address the question of whether vegetative incompatibility phenotype is associated with HET-c heteromeric and/or homomeric complex formation, the fusion construct het-c^OR::GFP was co-transformed with fusion constructs het-c^OR::HA or het-c^OR(PA)::HA into a het-c null strain. Commercially available anti-GFP and anti-HA antibodies were used in immunoprecipitation assay and Western blot analyses to detect the presence of HET-c^OR::GFP and HET-c::HA fusion proteins, respectively.
4.2 MATERIALS AND METHODS

4.2.1 Strains and media

*N. crassa* strains C2-2-9, C9-2, C15-1a, Xa-3 and CJ44 (see Chapter 3, Table 3-1) were used for this study. Genotypes and origins of these strains are described in Chapter 3 (Table 3-1). Vogel's (Vogel, 1964) synthetic media was used for cultures with required supplements. All cultures were maintained at 30°C.

4.2.2 Antibodies

Anti-GFP antibody was obtained from Boehringer Mannheim Co. (Cat. No. 1814460). It is a mixture of two mouse monoclonal antibodies (clone 7.1 and 13.1). The concentration of stock solution is 400 μg / ml. A 1:1000 dilution was optimized for Western blot analysis. A final concentration of 4 μg / ml was used for immunoprecipitation assays according to the manufacture's recommendations.

Rabbit anti-HA polyclonal antibody Y-11 was supplied by Santa Cruz Biotechnology Inc. (Cat. No. sc-805). The concentration of stock solution is 200 μg / ml. A 1:600 dilution was optimized for Western blot analysis. A final concentration of 5 μg / ml was used for immunoprecipitation assays according the manufacture’s recommendations.

Peroxidase conjugate anti-mouse IgG (whole molecule) was developed in rabbits and supplied by SIGMA (Lot. No. 018H4822, Cat. No. A9044). The working concentration for Western blot analysis was at a 1:8,000 dilution. Peroxidase conjugate anti-rabbit IgG (whole molecule) was developed in goat and also supplied by SIGMA Chemical Company (Cat. No.
The working concentration for Western blot analysis was optimized at a 1:5,000 dilution.

4.2.3 Construction of fusion genes

Recombinant DNA techniques and plasmid preparation were described as in Chapter 3 (Section 3.2.2). For the purpose of co-transformation, two N. crassa transformation vectors that have different selective markers, pCB1004 and pOKE 103 (see Chapter 3 Materials and Methods), were used for fusion gene constructions. The fusion gene het-c\textsuperscript{OR}::GFP was constructed in the vector pOKE103 by standard DNA manipulations (Sambrook et al., 1989). Fusion genes het-c\textsuperscript{OR}::HA and het-c\textsuperscript{OR(PA)}::HA were constructed by R. Todd in the vector pCB1004 as described below in section 4.2.3.2. All the fusion genes were designed to be expressed under the A. nidulans constitutive promoter, trpC. However, for the use of convenient restriction sites, all the fusion constructs contain the entire trpC and also part of the het-c promoter.

4.2.3.1 Fusion gene between het-c\textsuperscript{OR} and GFP

The GFP clone, SmGFP (soluble modified GFP, clone CD3-326), supplied by CLONTECH Laboratories Inc. (Palo Alto, California) was used in this study and GFP refers to SmGFP in the following text. Modifications of the SmGFP were described by Davis and Vierstra (1996). The het-c\textsuperscript{OR}::GFP fusion plasmid, pOKEpCOR::GFP, was constructed through several steps (Figure 4-1). First, the promoter trpC was cloned into pOKE103 at the Acc I / Xba I sites to generate plasmid pOKEpC. Second, an 800bp BamH I / Sac I fragment containing the entire GFP open reading frame was cloned into the plasmid pOKEpC to give
pOKEpCGFP. Third, a 2.7 kb Sca I/BamH I fragment containing approximately 160 bp of the promoter region and the C-terminal truncated open reading frame of het-cOR was cloned into pOKEpCGFP at the EcoR V/BamH I sites to give pOKEpCOR::GFP'. At this step, GFP was fused with the C-terminal of het-cOR at the BamH I site but was not in-frame. To make the het-cOR::GFP fusion in-frame, the plasmid pOKEpCOR::GFP' was digested with BamH I to generate a linear DNA fragment containing the BamH I overhang, CTAG, at both ends. An end-filling reaction with Klenow I DNA polymerase (Berger and Kimmel, 1987) at room temperature was performed for 1 hr to generate blunt ends of the linearized DNA. The reaction mixture was heat inactivated at 65° C for 10 min and ligated overnight to give plasmid pOKEpCOR::GFP in which both het-cOR and GFP were in frame. The DNA sequence of the fusion construct was confirmed by DNA sequencing (Appendix 7.3).
Figure 4-1. Outline of the \(\text{het-c}^{\text{OR}}::\text{GFP}\) fusion construction. The fusion construct contains both \(\text{trpC}\) and \(\text{het-c}\) promoters. Arrow in the sequences indicates \(\text{BamHI}\) cutting site. The \(\text{BamHI}\) overhang, CTAG, was filled with Klenow DNA polymerase I to generate blunt end double stranded DNA. Re-ligation of the \(\text{het-c}^{\text{OR}}::\text{GFP}\) resulted in an in-frame fusion between \(\text{het-c}^{\text{OR}}\) and \(\text{GFP}\).
4.2.3.2 Fusion genes between het-c and HA tag

The fusion genes het-c^{OR}::HA and het-c^{OR(PA)}::HA were made by Dr. R. Todd for the purpose of localizing HET-c (R. Todd, personal communication). A description of the construction of the plasmids is as follow. The het-c^{OR} was cloned into the Bluescript plasmid pKS at the Xba I / Sac I sites to generate the plasmid pKSCOR. The 27 nucleotide HA tag was inserted into het-c^{OR} prior to the stop codon by PCR mutagenesis. The plasmid pKSCOR was used as the template for PCR reaction. Two overlapping oligonucleotide primers named CHA1 and CHA2 were used. The sequences of the primers are: CHA1, 5'CTCGTAGAAGTATAACCTCGATGCTCGTCCG3' and CHA2, 5'TATTCGAATACGCGTAGCTTGGTGGAAGTAATG3'. The 5' end of the CHA1 oligonucleotide contains part of the HA tag sequences; the 5' of the CHA2 oligonucleotide contains the rest of the HA tag sequences, the het-c^{OR} stop codon, TGA, and the 3' end sequences of het-c^{OR}. The PCR reaction was performed according to the standard protocol as described in Chapter 3 (section 3.2.4) except that annealing was done at 50°C and pfu Taq polymerase (BRL, Bethesda research laboratory) was used to generate blunt end PCR products. The PCR products were ligated by standard protocol (Sambrook et al., 1989). The generated plasmid was called pKSCOR::HA (Figure 4-2). The DNA sequence of the fusion construct pKSCOR::HA was confirmed by DNA sequencing (R. Todd, personal communication). To generate the het-c^{OR(PA)}::HA construct, the approximately 2.44 kb Sca I / Kpn I fragment of het-c^{OR}::HA fusion gene that contains the het-c specificity region was replaced with the approximately 2.48 kb Sca I / Kpn I fragment from het-c^{PA} to generate the plasmid het-c^{OR(PA)}::HA. Since both EcoRV and Sca I generate blunt ends, the approximately 3.6 kb Sca I / Sac I fragment of het-c^{OR}::HA or het-c^{OR(PA)}::HA was subcloned into pCBpC at
EcoR V / Sac I sites to give plasmid \( pCBpCOR::HA \) or \( pCBpCOR(PA)::HA \), respectively (Figure 4-3).
Figure 4-2. Generating \( \text{het-}c^{\text{OR}} :: \text{HA} \) and \( \text{het-}c^{\text{OR(PA)}} :: \text{HA} \) fusion genes (constructed by R. Todd).
Figure 4-3. Cloning het-cOR::HA fusion gene into N. crassa expression vector pCB1004. The construct contains a trpC promoter and 126 bp of het-c promoter. Plasmid pCBpCOR(PA)::HA was constructed by the same principle (constructed by R. Todd).
4.2.4 Preparation of cell extracts

Approximately five compatible transformants or 20 – 30 incompatible \textit{het-c} transformants were transferred into Vogel’s selective liquid media containing hygromycin B and arginine and maintained at 30°C for three to four days. Mycelia from the cultures were harvested on Whatman No.1 filter paper through vacuum filtration. The following protocol was adapted from Peleg et al. (1996), with modifications. Approximately 0.5 g of mycelia was suspended in 2 ml of extraction buffer containing 20 mM Tris-HCl (pH 7.5), 50% (v/v) glycerol, 1 mM PMSF, 1 mM EDTA, 0.1 mM DTT and 0.1% Triton-100. Mycelia were ground with a Pyrex tissue grinder at 4°C until the mixture was homogenized. The homogenate was added to an equal volume of ice-cold acid-washed glass beads and incubated on ice for 30 seconds, followed by one minute vortex at high speed. The mixture was incubated on ice for 5 minutes. Cell debris was removed by centrifugation at 1,000 g until the supernatant was clear. The clear supernatant was the total soluble cell extract and was used for immunoprecipitation experiments.

4.2.5 Immunoprecipitation

Immunoprecipitation was performed using the Protein A Immunoprecipitation kit (Cat. No. 1 719 394, Boehringer Mannheim) following instructions in the manual with some optimization of the conditions. All incubation steps were performed at 4°C with gentle mixing. Five hundred µl of cell lysate was pre-absorbed with 3 µl of calf serum for 60 minutes and then incubated with 40 µl of protein A-Sepharose for an additional 60 minutes. This eliminates non-specific immuno-binding to protein A – Sepharose in the lysates. The mixture was centrifuged at 2,000 rpm for 30 seconds and the supernatant was transferred into
a new Eppendorf tube. Five to twelve µl of antibody (specifically, 12 µl of anti-HA polyclonal antibody Y-11 or 5 µl of pooled monoclonal anti-GFP antibody) were added to the supernatant and followed by overnight incubation at 4°C. Forty-five µl of protein A-Sepharose was then added to the mixture with a further incubation for 3 hours. The mixture was centrifuged at 2,000 rpm for 30 seconds and the supernatant was removed. The pellet was washed twice with low salt buffer B (0.2% NP-40, 10 mM Tris, pH 7.5, 0.15 M NaCl, 2 mM EDTA), one time with high salt buffer C (0.2% NP-40, 10 mM Tris, pH 7.5, 0.5 M NaCl, 2 mM EDTA), and finally with 10 mM Tris pH 7.5. During all washing steps, the mixture was collected by centrifugation at 2,000 rpm for 2 minutes. After the last wash, the residual washing buffer was removed carefully with a gauge needle. The pellet was resuspended in 40 µl of SDS sample buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 5% 2-mercaptoethanol and 0.25% bromophenol blue). Sepharose was pelleted from the suspension by centrifugation at 2,000 rpm for 1 minute and the supernatant was used for Western blot analysis.

4.2.6 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Immunoprecipitated samples were heated at 95°C for 3 minutes. Approximately 20 µl of each sample was subjected to SDS-PAGE according to modified Laemmli’s discontinuous buffer system (Laemmli, 1970; Laemmli and Favre, 1973) using a Bio-Rad Mini-Gel II apparatus (Bio-Rad). The electrophoresis conditions used were specified by the manufacturer. A 7.5% SDS-Acrylamide separating gel that allows effective separation of proteins between 200 and 40 kDa was used to separate samples containing the HET-c:GFP and HET-c:HA
fusion proteins. Broad range pre-stained molecular mass markers (BRL or NEB) were used as described by the supplier.

4.2.7 Western blotting

Transfer of protein to nitrocellulose membranes from Laemmli SDS-PAGE was accomplished using the Towbin buffer system (Towbin et al., 1979) and either the Bio-Rad Trans-Blot™ or Mini Trans-Blot™ electrophoretic transfer apparatus (Bio-Rad). Transfer was accomplished at 30 V overnight. Nitrocellulose membranes (Protein™, Schleicher & Schuell) were first hydrated in distilled water and then equilibrated in Towbin's transfer buffer for at least 15 minutes before preparing the gel/membrane transfer sandwich. The efficiency of transfer was monitored by Coomassie Blue staining of the gel. The membrane with transferred proteins was rinsed twice with TBS buffer (50 mM Tris, pH 7.5, 150 mM NaCl) and then incubated with 3% BSA in TBS at room temperature with gentle agitating. After 1 hour, the membrane was transferred into primary antibody solution and further incubated for 1 hour. The membrane was washed three times in TBST (TBS plus 0.1% Tween 20) and once in 1% BSA-TBS for 10 minutes each. The membrane was then transferred into secondary antibody solution for 30 minutes at room temperature with agitation. After the membrane was washed four times with TBST for 15 minutes each, specific protein bands were visualized using a BM Chemiluminescence kit (Boehringer Mannheim) or ECL™ kit (RPN 2106, Amersham Pharmacia Biotech., UK) and Kodak X-Omat film.
4.3 RESULTS

4.3.1 Construction of fusion genes

The predicted polypeptides resulting from the fusion gene constructs are shown in Figure 4-4. Fusion constructs \( pCBpCOR::HA \), \( pCBpCOR(PA)::HA \) (both made by R. Todd) and \( pOKEpCOR::GFP \) were made at independent experimental times for the purpose of localizing HET-c.

Mutational analysis showed that the truncation of the HET-c glycine rich C-terminus did not affect \( \text{het-c} \) incompatibility by transformation assays (Saupe et al., 1996b, and R. Todd, personal communication). In the construction of \( \text{het-c}^{OR}::GFP \) fusion genes, the \( \text{BamHI} \) site in the 3'-end of the \( \text{het-c}^{OR} \) and 5'end of the \( \text{GFP} \) gene was used for generating recombinant molecules. The HET-c\(^{OR}::GFP \) fusion protein contains 714 amino acids of HET-c\(^{OR} \), 266 amino acids of GFP, with an additional six amino acids in-between that are from the end-filling of the \( \text{BamHI} \) overhang, CTAG, and partial DNA sequences of the \( \text{GFP} \) promoter (Figure 4-1). The fusion protein is predicted to have a molecular weight of approximately 128 kDa (Figure 4-4A). The fusion protein HET-c\(^{OR}::HA \) contains the entire 966 amino acids of HET-c\(^{OR} \) plus nine amino acid residues of the HA epitope and has a predicted molecular weight of approximately 127 kDa (Figure 4-4B). The fusion protein HET-c\(^{OR(PA)}::HA \) contains 730 amino acids of HET-c\(^{PA} \) at the N-terminal and 255 amino acids of HET-c\(^{OR} \) at the C-terminal plus nine amino acid residues of the HA epitope and has a predicted molecular weight of approximately 128 kDa (Figure 4-4C).
A. HET-c\textsuperscript{OR}:GFP  982 aa  128 kDa

B. HET-c\textsuperscript{OR}:HA  975 aa  127 kDa

C. HET-c\textsuperscript{OR(PA)}:HA  985 aa  128 kDa

Figure 4-4. Summary of the predicted fusion gene products.
4.3.2 *het-c* activity of the fusion genes

To confirm that the *het-c* fusion genes retain the capacity to mediate vegetative incompatibility and thus presumably would function like wild type *het-c*, fusion gene constructs were introduced into strains with different *het-c* specificities to assay incompatibility. The phenotypes of the transformants are summarized in Table 4-1. Introduction of fusion gene constructs pCBpCOR::HA or pOKEpCOR::GFP into the *het-c*<sup>PA</sup> strains C9-2 or Xa-3 (see Chapter 3, Table 3-1) gave class III transformants that showed the same phenotype as introduction of *het-c*<sup>OR</sup> into a *het-c*<sup>PA</sup> strain. The transformants displayed severely inhibited growth, curly, aconidial colony morphology and occurrence of dead hyphal compartments. Introduction of the fusion gene construct pCBpCOR(PA)::HA into the *het-c*<sup>OR</sup> strain C2-2-9 (see Chapter 3, Table 3-1) gave class II (intermediate incompatibility) transformants that showed the same phenotype as introduction of *het-c*<sup>PA</sup> into a *het-c*<sup>OR</sup> strain. The transformants displayed inhibited growth, abnormal hyphal morphology and occurrence of dead hyphal compartments after two days of growth (data not shown). Only vigorous class I (compatible) transformants were obtained when fusion gene constructs pCBpCOR:HA or pOKEpCOR::GFP were introduced into the *het-c*<sup>OR</sup> strains C2-2-9 or C15-1a and when the fusion gene construct pCBpCOR(PA)::HA was introduced into a *het-c*<sup>PA</sup> strain C9-2. The class I transformants had the identical phenotype as introduction of *het-c*<sup>OR</sup> into a *het-c*<sup>OR</sup> strain or introduction of *het-c*<sup>PA</sup> into a *het-c*<sup>PA</sup> strain, and displayed vigorous growth and no occurrence of dead hyphal compartments. These transformation results indicate that the *het-c* fusion gene constructs had retained wild type *het-c* activity and presumably mediated vegetative incompatibility in an identical manner to non-tagged *het-c*.
Table 4-1. Summary of the phenotypes of transformants containing fusion gene constructs.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Phenotype of transformants when the construct was introduced into different recipient strains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C2-2-9 (het-c&lt;sup&gt;OR&lt;/sup&gt;)</td>
</tr>
<tr>
<td>pCB1004</td>
<td>++</td>
</tr>
<tr>
<td>pOKE103</td>
<td>ND</td>
</tr>
<tr>
<td>het-c&lt;sup&gt;OR&lt;/sup&gt;-pCB1004</td>
<td>++</td>
</tr>
<tr>
<td>het-c&lt;sup&gt;PA&lt;/sup&gt;-pCB1004</td>
<td>-</td>
</tr>
<tr>
<td>het-c&lt;sup&gt;OR&lt;/sup&gt;-pOKE103</td>
<td>ND</td>
</tr>
<tr>
<td>pCBpCOR::HA</td>
<td>++</td>
</tr>
<tr>
<td>pCBpCOR(PA)::HA</td>
<td>-</td>
</tr>
<tr>
<td>pOKEpCOR::GFP</td>
<td>ND</td>
</tr>
</tbody>
</table>

Phenotypes of ++, - and -- type of transformants were described in Chapter 3, Table 3-5. ND = not determined.

4.3.3 Heteromeric complex formation of HET-c proteins

To examine if HET-c proteins form stable heteromeric complexes during vegetative incompatibility and/or stable homomeric complexes during vegetative growth, the fusion gene construct pOKEpCOR::GFP was co-transformed with pCBpCOR(PA)::HA or pCBpCOR::HA into the het-c deletion strain CJ44 (het-c<sup>N</sup>; pan-2; arg-5) by co-selection of hyg<sup>R</sup> and PAN requirements. For the negative control experiments, het-c<sup>OR</sup> (in vector pOKE103) was co-transformed with het-c<sup>OR</sup>, het-c<sup>PA</sup> or het-c<sup>OR(PA)</sup>::HA (all in vector pCB1004) (see Chapter 3, Material and Methods), and the transformants were co-selected for hyg<sup>R</sup> and PAN requirements. An equal amount of cell lysates of these transformants was analyzed by two independent immunoprecipitation assays. To gain sufficient cell lysates for analyses,
pooled transformants (five compatible transformants or 20 incompatible transformants) were used.

First, pooled (mixed) anti-GFP antibody was added to the total cell lysate for binding to Protein A-Sepharose. To ensure the efficiency of immuno-binding and immuno-reaction, immunoprecipitation samples were analysed redundantly with anti-GFP antibody by Western blotting (Figure 4-5). Equal amounts of protein of immunoprecipitated samples from five sets of co-transformants, het-c^{OR}:GFP/het-c^{OR(PA)}::HA, het-c^{OR}:GFP/het-c^{OR}:HA, het-c^{OR}/het-c^{PA}, het-c^{OR}/het-c^{OR(PA)}::HA, het-c^{OR}/het-c^{OR}, were analyzed (the staining gel is not shown). A highly abundant band of approximately 55 kDa that corresponds to the molecular weight of the heavy chain of IgG is present in all samples. The presence of the corresponding Ig G band (55 kDa) confirmed the binding efficiency of the antibody to Protein A. The nearly identical density of the 55 kDa band in all the lanes indicates that a nearly identical quantity of anti-GFP antibody was bound to protein A in all immunoprecipitation samples. Three bands of approximately 126 kDa, 98 kDa and 90 kDa are observed only in the transformants that contain the het-c^{OR}:GFP fusion construct (lanes 1 and 2 in Figure 4-5). The 126 kDa band corresponds to the molecular weight of the predicted HET-c^{OR}:GFP fusion protein. The 98 kDa and the 90 kDa bands could be the degradation products of the HET-c^{OR}:GFP fusion protein (see 4.4 Discussion). The bands of 92 kDa to 96 kDa are present non-specifically in all samples with comparable density (see comments in next paragraph).

To determine if HET-c proteins form stable heteromeric or homomeric complexes during vegetative growth, equal amounts of samples from the anti-GFP antibody immunoprecipitation assays were analyzed with rabbit anti-HA polyclonal antibody (Y-11) by Western blotting (the staining gel is not shown). As shown in Figure 4-6, two bands of
approximately 127 kDa and 98 kDa are present specifically in the sample from transformants containing the \textit{het-c}^{OR::GFP} and \textit{het-c}^{OR(PA)::HA} constructs (lane 1 in Figure 4-6). The 127 kDa band corresponds to the molecular weight of the predicted \textit{HET-c}^{OR(PA)::HA} fusion protein. The 98 kDa band may be the N-terminal degradation product of the \textit{HET-c}^{OR(PA)::HA} fusion protein. The presence of the 98 kDa band will be discussed in Section 4.4. Two bands corresponding to 92 - 96 kDa are also present non-specifically in all the samples with comparable amount in this analysis (labeled NS). The presence of these non-specific bands may be due to contamination of non-specific proteins (such as protein-A - like proteins) in the immunoprecipitation samples that bind non-specifically to IgG. These results suggested that \textit{HET-c} proteins (both the entire protein and the degradation product) are likely to form a heteromeric complex only in the incompatible transformants that contained \textit{het-c}^{OR::GFP} and \textit{het-c}^{OR(PA)::HA} fusion constructs and that no homomeric complex was detected in transformants containing \textit{het-c}^{OR::GFP} and \textit{het-c}^{OR::HA} fusion constructs.

The observation of \textit{HET-c} heteromeric complex formation in incompatible transformants was confirmed independently by immunoprecipitation assays with rabbit anti-HA antibody (Y-11). Samples from the same five sets of co-transformants, \textit{het-c}^{OR::GFP/het-c}^{PA(OR)::HA}, \textit{het-c}^{OR::GFP/het-c}^{OR::HA}, \textit{het-c}^{OR/het-c}^{PA}, \textit{het-c}^{OR/het-c}^{PA(OR)::HA}, \textit{het-c}^{OR/het-c}^{OR}, were used for this assay. As described above, to ensure the efficiency of the immunoprecipitation reaction, equal amounts of proteins of all samples were first analyzed redundantly with anti-HA (Y-11) antibody by Western blotting (Figure 4-7; the staining gel is not shown). Two bands of approximately 55 kDa and 79 kDa are present in all samples. The 55 kDa band present in all samples corresponds to the molecular weight of the IgG heavy chain. The nearly equal density of the 55 kDa bands in all samples indicating equal quantity
of anti-HA antibody was bound to the protein A in all immunoprecipitation reactions. The 79 kDa band present in all the samples with nearly identical density could be the aggregation product of IgG heavy chains and light chains (55 kDa + 25 kDa, the staining gel is not shown). Two bands of approximately 127 kDa and 98 kDa are observed specifically and nearly equal in transformants that contain het-c::HA fusion constructs (lanes 1, 2 and 4 in Figure 4-7). The 127 kDa band present in lanes 1, 2 and 4 in Figure 4-7 corresponds to the molecular weight of the predicted HET-c::HA fusion proteins. The presence of the 98 kDa band will be discussed in section 4.4.

The anti-HA antibody immunoprecipitation samples were further analyzed by Western blotting with mouse anti-GFP antibody. As presented in Figure 4-8, two bands of approximately 127 kDa and 98 kDa were present specifically in samples from incompatible transformants that contained het-cOR::GFP and het-cOR(PA)::HA constructs (lane 1 in Figure 4-8). No specific band was observed in the transformants that contain het-cOR::GFP and het-cOR::HA constructs (lane 2 in Figure 4-8). The 127 kDa band correspondings the predicted molecular weight of HET-cOR::GFP fusion protein. As previously mentioned, the 98 kDa band is likely the N-terminal truncation product of the HET-cOR::GFP fusion protein. The presence of the 98 kDa band will be discussed in section 4.4. These data are consistent with the observation in the immunoprecipitation assay with anti-GFP antibody that HET-c proteins form a heteromeric complex during vegetative incompatibility.
Figure 4-5. Western blot analysis of mouse anti-GFP immunoprecipitation samples with mouse anti-GFP antibody. Lanes 1 – 5 represent transformants containing different *het-c* constructs. 1, *het-c<sup>OR</sup>::GFP/het-c<sup>OR(PA)</sup>::HA. 2, *het-c<sup>OR</sup>::GFP/het-c<sup>OR</sup>::HA. 3, *het-c<sup>OR</sup>/het-c<sup>PA</sup>. 4, *het-c<sup>OR</sup>/het-c<sup>OR(PA)</sup>::HA. 5, *het-c<sup>OR</sup>/het-c<sup>OR</sup>. Numbers on the right indicate the marker for molecular weight (kDa). Linear arrow indicates the band (approximately 126 kDa) corresponding to the predicted molecular weight of HET-c<sup>OR</sup>::GFP fusion protein. Block arrow indicates the 55 kDa IgG heavy chain. NS, nonspecific proteins binding nonspecifically to IgG. The lower bands in lane 1 and lane 2 (about 98 kDa and 90 kDa) may be the degradation products of the HET-c<sup>OR</sup>::GFP fusion protein (see discussion).
Figure 4-6. Western blot analysis of mouse anti-GFP immunoprecipitation samples with polyclonal anti-HA antibody (Y-11). Lanes 1 – 5 represents transformants containing different het-c constructs. 1, het-c<sup>OR</sup> :: GFP/het-c<sup>OR(PA)</sup> :: HA. 2, het-c<sup>OR</sup> :: GFP/het-c<sup>OR</sup> :: HA. 3, het-c<sup>OR</sup> /het-c<sup>PA</sup>. 4, het-c<sup>OR</sup> /het-c<sup>OR(PA)</sup> :: HA. 5, het-c<sup>OR</sup> /het-c<sup>OR</sup>. Numbers on the right indicate the marker for molecular weight (kDa). Linear arrow indicates the band (approximately 127 kDa) corresponding to the predicted molecular weight of HET-c::HA fusion protein. NS, nonspecific proteins binding non-specifically to IgG. The lower bands in lane 1 (about 98 kDa) may be the degradation products of the HET-c::HA fusion protein (see discussion).
Figure 4-7. Western blot analysis of rabbit anti-HA (Y-11) immunoprecipitation samples with anti-HA antibody (Y-11). Lanes 1 – 5 represents transformants containing different het-c constructs. 1, \( \text{het-c}^{\text{OR}} \cdot \text{GFP/}\text{het-c}^{\text{OR(PA)}} \cdot \text{HA} \). 2, \( \text{het-c}^{\text{OR}} \cdot \text{GFP/}\text{het-c}^{\text{OR}} \cdot \text{HA} \). 3, \( \text{het-c}^{\text{OR}} / \text{het-c}^{\text{PA}} \). 4, \( \text{het-c}^{\text{OR}} / \text{het-c}^{\text{OR(PA)}} \cdot \text{HA} \). 5, \( \text{het-c}^{\text{OR}} / \text{het-c}^{\text{OR}} \). Numbers on the right indicate the marker for molecular weight (kDa). Linear arrow indicates the band (approximately 127 kDa) corresponding to the predicted molecular weight of HET-c:HA fusion protein. Block arrow indicates the 55 kDa IgG heavy chain. NS, nonspecific bands (approximately 79 kDa) or possibly the aggregation product of IgG heavy chain and light chain. The lower band in lane 1, 2 and 4 (about 98 kDa) may be the degradation products of the HET-c::HA fusion protein (see discussion).
Figure 4-8. Western blot analysis of rabbit anti-HA (Y-11) immunoprecipitation samples with anti-GFP antibody. Lanes 1 – 5 represents transformants containing different het-c constructs. 1, het-c_{OR}::GFP/het-c_{OR(PA)}::HA. 2, het-c_{OR}::GFP/het-c_{OR}::HA. 3, het-c_{OR}/het-c_{PA}. 4, het-c_{OR}/het-c_{OR(PA)}::HA. 5, het-c_{OR}/het-c_{OR}. Numbers on the right indicate the marker for molecular weight (kDa). Linear arrow indicates the band (approximately 127 kDa) corresponding to the predicted molecular weight of HET-c::GFP fusion protein. The lower band in lane 1 (about 98 kDa) may be the degradation products of the HET-c_{OR}::GFP fusion protein (see discussion).
4.4 DISCUSSION

4.4.1 Non-self recognition is mediated via protein heteromeric complex formation

Two independent immunoprecipitation assays were performed in this study with consistent results. Both support the hypothesis that HET-c proteins form a heteromeric complex during vegetative incompatibility and that a stable homomeric complex is not formed during vegetative growth.

The first immunoprecipitation used anti-GFP antibody (Figure 4-5 and Figure 4-6) and the second immunoprecipitation used anti-HA antibody (Figure 4-7 and Figure 4-8). In the first redundant Western blot (Figure 4-5), a band of approximately 126 kDa that corresponds to the predicted molecular weight of the HET-c\textsuperscript{OR}::GFP fusion protein and two bands of approximately 98 kDa and 90 kDa were present only in transformants containing the \textit{het-c}\textsuperscript{OR}::GFP fusion construct (lanes 1 and 2 in Figure 4-5). In the second redundant Western blot (Figure 4-7), a band of approximately 127 kDa that corresponds to the predicted molecular weight of HET-c\textsuperscript{OR}::HA and HET-c\textsuperscript{OR(PA)}::HA fusion proteins and a band of approximately 98 kDa were present only in transformants containing \textit{het-c}\textsuperscript{OR}::HA and \textit{het-c}\textsuperscript{OR(PA)}::HA fusion constructs (lanes 1, 2 and 4 in Figure 4-7). These data suggest that the 126-127 kDa band is likely the specific band of HET-c fusion proteins and that the 98 kDa and the 90 kDa bands are probably degradation products of the fusion proteins.

In both Western blots analyses with alternative antibodies (anti-HA antibody in the first assay and anti-GFP antibody in the second assay, Figure 4-6 and Figure 4-8), two specific bands of approximately 127 kDa and 98 kDa were present only in transformants containing both \textit{het-c}\textsuperscript{OR}::GFP and \textit{het-c}\textsuperscript{OR(PA)}::HA constructs (lane 1 in Figures 4-6 and 4-8).
These results indicate that HET-c^{OR} and HET-c^{PA} can form a heteromeric complex, and that a HET-c^{OR} and HET-c^{OR} homomeric complex is not detectable. These data support the hypothesis that HET-c heteromeric complex formation is the mechanism mediating non-self recognition, although the dynamics of HET-c heteromeric complex formation are not clear.

In Western blots, putative HET-c fusion protein degradation products were observed. A specific band of approximately 90 kDa occurred in samples containing het-c^{OR}::GFP fusion constructs in Figure 4-5 (lanes 1 and 2). It could be inferred that the approximate 90 kDa band is probably the N-terminal degradation product of the HET-c^{OR}::GFP fusion protein due to sample manipulation, since the corresponding band did not appear in other blots. A band of approximately 98 kDa was present in all the blots (both compatible and incompatible transformants) containing het-c^{OR}::GFP and/or het-c::HA fusion constructs. Since it is consistently present in all blots and all transformants containing het-c fusion constructs, it is probable that the 98 kDa protein is the degradation product of the HET-c fusion protein. Since both HA and GFP were fused with HET-c at the C-terminal, it is probable that the N-termini of the fusion proteins was truncated. This truncation apparently occurs at the same site in the HET-c^{OR}::GFP and HET-c^{OR(PA)}::HA and is approximately 220-230 amino acids distant from the N-terminal. In this case, however, the 98 kDa protein would retain the specificity region of HET-c (see Chapter 1, Figure 1-7). The 98 kDa band is present in the same pattern as the predicted HET-c fusion proteins (shown as 126 and 128 kDa in the blots), which indicates that a heteromeric complex was also formed between the 98 kDa products (truncated HET-c^{OR}::GFP and HET-c^{OR(PA)}::HA). However, no differences were observed in the presence of the 98 kDa band between compatible and incompatible transformants, suggesting that the protein truncation is not specific to the vegetative
incompatibility reaction. Possible explanations for this observation could be: 1) a secondary cellular function of the truncated HET-c protein; and 2) posttranslational modifications of HET-c polypeptides.

The HET-c peptide is predicted to have a glycine rich C-terminal domain that is similar to glycine rich domains found in a number of extracellular or cell envelope proteins (Saupe et al., 1996b). It is possible that HET-c might be localized in the cell wall or could have a secondary cellular function as a structural protein of the cell wall components. However, microinjection experiments conducted by Wilson et al. (1961) indicated that het-c vegetative incompatibility in *N. crassa* was mediated by cytoplasmic elements. The capability of heterokaryon formation indicated that strains with different het-c genetic background are capable of cell fusion although the fusion cells might not be viable. These observations suggest that HET-c within the cell wall may not be important in triggering vegetative incompatibility reaction. Due to technical difficulties, we were unable to localize HET-c protein (R. Todd, personal communication and this study). However, we can not rule out the possibility of dual functions of HET-c. First, HET-c functions as a cell wall component, in which case heteromeric complex formation mediates non-self recognition and thus triggers the formation of a specific cell wall structure, such as a septal plug, during vegetative incompatibility. Second, HET-c functions as a cytoplasmic component, in which case heteromeric complex formation also mediates non-self recognition but triggers cellular responses, such as cell organelle and nuclear DNA degradation during vegetative incompatibility.

The dual functions, cell wall component and cytoplasmic component, may require different forms of HET-c polypeptides: N-terminal truncated (98 kDa) and non-modified
(~127 kDa) forms. Since the N-terminal truncation would not affect the specificity region and
the leucine zipper motif, both forms would presumably be able to form heteromeric
complexes during vegetative incompatibility. Due to the technical aspects in this study, it is
likely that the cell wall components in the cell lysates may be mostly eliminated. Also due to
the fact that cellular response of vegetative incompatibility does not require a high expression
level of HET-c (Saupe and Wu, personal observations), whether truncated HET-c is localized
in the cell wall or cytoplasm could not be distinguished by Western blot analysis in this
study. Whether the 98 kDa product is due to the N-terminal truncation of the fusion protein
could be determined by N-terminal sequencing, and whether the truncated HET-c is a cell
wall component could be determined by a combination of HET-c N-terminal deletion and
localization studies.

In summary, this study supports the hypothesis that non-self recognition during
vegetative incompatibility is mediated via HET-c heteromeric complex formation. These data
also indicate that HET-c^{OR} does not form a homomeric complex in the soluble component of
the cytoplasm. Due to the elimination of the cell wall components in the cell lysates in this
study, it can not be excluded that HET-c^{OR} may form stable homomeric complex in the cell
wall as a structural component.
4.4.2 Possible mechanisms for heteromeric complex formation

The predicted HET-c peptides contain a leucine rich heptad repeat motif with hydrophobic residues at positions a and d (Figure 4-9; Saupe et al., 1996b; Saupe and Glass, 1997). This motif putatively forms amphipathic \( \alpha \)-helices that are found as a dimerization motif (termed leucine-zipper or coiled-coil structure) in several transcriptional regulators and fibrous structural proteins (O’shea et al., 1989b, 1991, 1992; Lehrer and Stafford, 1991). Mutational analysis has shown that an in-frame deletion of the leucine rich haptad repeat in \( het-c \) abolished its ability to mediate vegetative incompatibility (Saupe et al., 1996b). It is possible that the leucine rich coiled-coil motif in HET-c is essential in mediating HET-c heterocomplex formation.

The 34-48 amino acid variable region in HET-c was found to be sufficient for conferring allelic specificity (Saupe and Glass, 1997). Introduction of alleles \( pdl \) and \( pd2 \) that have alterations in the size of the PA-insertion motif into a \( het-c^{OR} \) strain resulted in class III and class II transformants, respectively, which may be due to different heteromeric complexes formed between PD1-HET-c\(^{PA} \) and PD2-HET-c\(^{PA} \) (Chapter 3). Immunoprecipitation data indicated that different HET-c proteins formed a heterocomplex, while the same HET-c proteins did not form a homocomplex. These studies suggest that the specificity region of HET-c may also be important in mediating heterocomplex formation and non-self recognition.
Leucine zipper coiled-coil motif mediating protein heterodimerization has been demonstrated in several systems. A well understood example of protein heterodimer formation is the Fos–Jun protein interactions in which preferential heterodimer formation by the Fos and Jun leucine zipper regions is largely a thermodynamic consequence of Fos homodimer instability (O’Shea et al., 1992). Fos and Jun, which are the protein products of the nuclear proto-oncogenes c-fos and c-jun, bind to DNA as dimers and modulate transcription of a wide variety of genes in response to mitogenic stimuli (Ransone and Verma, 1990). The Jun protein forms a stable homodimer that can bind to DNA, while the Fos protein does not form a stable homodimer to bind DNA. However, the Fos-Jun proteins preferentially form a heterodimer over homodimer thermodynamically (O’Shea et al., 1992).
Both Fos and Jun contain a single leucine zipper region that is necessary and sufficient to mediate preferential heterodimer formation. The leucine zipper regions from Fos and Jun are known to fold as two-stranded, parallel coiled-coil helices (O’Shea et al., 1989; 1991) that prefer to interact with each other rather than with themselves (O’Shea et al., 1992). Eight amino acids from the Fos and Jun proteins were identified that were sufficient to mediate specificity and preferential heterodimer formation, while acidic residues in the Fos leucine zipper region were found to destabilize the Fos homodimer formation through thermodynamic interactions (O’Shea et al., 1992). In *S. cerevisiae*, the homeodomain proteins α1 and α2 form heterodimers to repress haploid-specific genes in diploid cells and allow post-fertilization development, while α2 homodimers repress a-specific genes. α1 does not form homodimers (Herskowitz, 1989; Ho et al., 1994). Ho et al. (1994) identified 3, 4-hydrophobic heptad repeats within the N-terminus of the α1 and α2 proteins and proposed that these motifs mediate dimerization by two leucine zipper-like coiled-coils.

Crystallographic studies of the α1/α2 heterodimer have revealed that hydrophobic interactions and several hydrogen bonds between the coiled-coil interface of α1 and α2 proteins primarily promote α1/α2 heterodimerization and discourage α1/α1 homodimerization (Li et al., 1995). Leucine zippers mediating heterodimer formation was also shown in the yeast transcriptional activator GCN4 (O’Shea et al., 1989b; 1991). Destabilization of a homodimer were also shown in the case of the tropomyosin αβ heterodimerization (Lehrer et al., 1989; Lehrer and Stafford, 1991).

A specificity region functioning in heteromeric complex formation and mediating non-self recognition has been described in several fungal mating systems. In *U. maydis*, the b mating-type genes have variable and constant regions (Kronstad and Leong, 1990). Yeast
two-hybrid and immunoprecipitation studies indicated that the N-terminal variable regions of the bE and bW proteins determined the specificity and also functioned in bE-bW heterodimerization and that interactions between the bE and bW proteins from the same allele did not occur (Kamper et al., 1995). In C. cinereus, specificity determinants of the A mating-type proteins HD1 and HD2 were localized to the N-terminal 160-170 amino acids that were also important for HD1-HD2 heterodimer formation (Banham et al., 1995). In S. commune, the specificity region of the Y4 and Z5 proteins were found playing an important role in both Y-Z recognition and heteromeric complex formations, while interactions between Y4 and Z4 (a self combination) did not occur (Wu et al., 1996; Yue et al., 1997).

The data presented in this study and protein secondary structure prediction (Chapter 3) plus mutational analyses (Saupe et al., 1996b) suggested a possible mechanism for mediating HET-c heterocomplex formation (Figure 4-10). The model hypothesizes that the leucine rich coiled-coil motif provides sufficient general contacts to form HET-c protein complexes, perhaps via hydrophobic interactions. No preference between a homomeric complex and a heteromeric complex formation occurs, since the coiled-coil motif is very similar among HET-c^{OR}, HET-c^{PA} and HET-c^{GR}. However, the stability of the complex may be regulated by the conformations of the specificity domain of the HET-c proteins. Different conformations in the specificity domain between alternative HET-c proteins may promote the dimerization to form heterodimer; identical or similar conformations in the specificity domain between two HET-c proteins may destabilize the dimerization and thus prevent homodimer formation. The stability of the complex may be due to thermodynamics that are primarily contributed by hydrophobic interactions and/or polar interactions between HET-c proteins or interactions with other proteins.
The region of HET-c that mediates protein-protein interactions has not been defined in this study. Structural and genetic studies (e.g. yeast two-hybrid assays) are required to examine whether the leucine rich coiled-coil motif and the specificity domain are directly involved in HET-c heterocomplex formation. It is possible that a specific region is not directly involved in protein-protein interactions and that it only regulates the equilibrium of the protein complex formation. It is also possible that HET-c proteins do not directly interact at all; instead, they may interact with a third protein (a docking protein) that is required to form a HET-c heterocomplex. However, these possibilities can now be tested experimentally.
Figure 4-10. Schematic model for stable HET-c heterocomplex formation. The leucine zipper motif (indicated by L in open box) may provide general interactions for protein complex formation. The conformation of the HET-c specificity region (indicated by different filled blocks for allele-specific proteins) may regulate the stability of the protein complex, possibly via hydrophobic thermodynamics. A different conformation may promote heterocomplex formation; the same conformation may discourage homocomplex formation.
Chapter 5  Concluding Remarks

5.1  Summary

To understand the biological role of vegetative incompatibility in fungal populations and the molecular and biochemical mechanisms of \textit{het-c} mediated vegetative incompatibility, three fundamental questions were addressed in this thesis. First, is the \textit{het-c} locus subject to balancing selection for the role of non-self recognition? Second, what is the primary determinant for \textit{het-c} allelic specificity? Third, is the non-self recognition mediated by the mechanism of protein heteromeric complex formation?

The \textit{het-c} locus displayed trans-specific evolution and long persistence of polymorphisms within the Sordariaceae. These data indicated that the \textit{het-c} locus is under balancing selection presumably to mediate non-self recognition during vegetative growth. An increased number and frequency of non-synonymous nucleotide substitutions in the \textit{het-c} polymorphic domain revealed positive selection favoring the diversity of HET-c peptide for its role in mediating non-self recognition. This study supports the hypothesis that vegetative incompatibility is the \textit{bona fide} non-self recognition system in filamentous ascomycetes in the Sordariaceae, and perhaps in other filamentous fungi as well. Comparable evidence for the persistence of ancient polymorphisms at the \textit{het-c} locus, mammalian \textit{MHC} loci and plant \textit{S} loci reinforces the view that balancing selection is an evolutionary mechanism common to different non-self recognition systems.

The isolation of the specificity region from a large number of naturally occurring \textit{het-c} alleles enabled me to investigate whether amino acid substitution in the specificity
domain affected het-c allelic specificity. Chimeric construction and transformation assays indicated that substitutions of amino acid residues in the variable region do not affect het-c allelic specificity. Construction of artificial het-c alleles and transformation assays revealed that the pattern of insertion and the size of the PA-specific insertion are the primary determinants for het-c allelic specificity. Based on the data presented in this study and protein secondary structure prediction, a HET-c protein heterodimerization model was proposed for mediating recognition during vegetative incompatibility.

The immunoprecipitation assay results indicated that HET-c\textsuperscript{OR} and HET-c\textsuperscript{PA} form a heteromeric complex during vegetative incompatibility while a HET-c\textsuperscript{OR} homomeric complex is not formed (at least not in the soluble component) during vegetative growth. The data suggested that non-self recognition is mediated via the mechanism of HET-c heteromeric complex formation during vegetative incompatibility.

5.2 General model of how het-c mediates vegetative incompatibility

Vegetative incompatibility reactions appear to have conserved features in filamentous fungi. In \textit{N. crassa} and \textit{P. anserina}, the vegetative incompatibility response involves common stages such as hyphal compartmentation, vacuolization and death (Begueret et al., 1994; Jacobson et al., 1998; N. L. Glass, unpublished results), suggesting that common mechanisms mediate vegetative incompatibility. In \textit{P. anserina}, a “poison heteromeric complex” model, in which heteromeric complexes between the products of incompatible genes are lethal to the cell, has been proposed for mediating vegetative
incompatibility based on genetic and mutational studies (Begueret et al., 1994). However, no direct evidence is available to support this hypothesis.

In N. crassa, mutational and genetic analysis indicated that the vegetative incompatibility response requires the co-existence of two incompatible het-c alleles (Saupe et al., 1996; Xiang, unpublished results). A strain that has a mutated het-c allele can not display vegetative incompatibility (Saupe et al., 1996b). Immunoprecipitation results presented in this study suggested that het-c mediates recognition during vegetative incompatibility via the formation of heteromeric complex of different het-c products (Chapter 4). Chimeric het-c allele construction and secondary structure prediction suggested that the specificity domain of het-c may be involved in the mechanism of stable HET-c heteromeric complex formation. The stability of the heteromeric complex may be regulated by the specificity domains from two alternative HET-c proteins, which may be associated with the observation of different phenotypes in incompatible transformants (Chapters 3 and 4).

Observations of the phenotype variations in vegetative incompatibility (Jacobson et al., 1998; chapter 3) indicated that incompatibility response mediated by het-c may be via different mechanisms. First, as proposed in P. anserina, the HET-c heteromeric complex may directly “poison” the cell; it may have a direct negative affect on cell growth. A threshold level of the heteromeric complex may be required to have a lethal effect on cells and thus trigger hyphal compartmentation and death (Figure 5-1). If formation of the heteromeric complex does not reach the threshold level, hyphae may display different incompatibility phenotypes. Second, the active HET-c heteromeric complex may induce sequential cellular events and thus trigger the incompatibility
response (Figure 5-1). The heteromeric complex may induce incompatibility reaction via interacting with a third molecule (X) or via regulating the expression or activity of the molecule (Y) that is essential in regulating cell growth (or death).

Indirect evidence exists to support this model. First, morphological and biochemical studies indicated that het-c mediated vegetative incompatibility has common features with programmed cell death (PCD) in animal and plant cells (Jacobson et al., 1998; Marek et al., 1999). Secondly, the severity of incompatibility reaction was different in het-c reciprocal transformants either between het-c^{OR} and het-c^{PA} or between het-c^{GR} and het-c^{PA} (Chapter 3). The different incompatible phenotypes could be due to different levels of heteromeric complex formation, interactions of the heteromeric complex with different modulators (X) in the recipient strains or heteromeric complex regulating the expression or activity of different targeting molecules (Y) (Figure 5-1).
Figure 5-1. General model for *het-c* mediated vegetative incompatibility. The model depicts the heteromeric complex formation of the products of incompatible *het-c* alleles. The HET-c heteromeric complex may trigger vegetative incompatibility by a lethal affect on cell growth, interacting with the modulator X, or regulating the activity or expression of the Y that may induce vegetative incompatibility (see the text for details). OR, HET-c^OR^. PA, HET-c^PA^. 
5.2 Future directions

The biological significance of vegetative incompatibility and the molecular basis of \textit{het-c} allelic specificity have been examined in this study. HET-c heteromeric complex formation is shown to be a mechanism for mediating non-self recognition during vegetative incompatibility based on the data of the immunoprecipitation assay. However, many questions remain to be further addressed.

The \textit{het-c} locus in the Sordariaceae displayed trans-specific polymorphism and is shown to be under balancing selection. Is \textit{het-c} conserved and does the \textit{het-c} locus also function as a \textit{het} locus in species other than in the Sordariaceae? Chimeric construction examined the molecular basis for \textit{het-c} allelic specificity determination. However, whether the specificity domain is directly involved in recognition mediated by heteromeric complex formation and how differences in the specificity domain affect heteromeric complex formation need to be investigated by yeast two-hybrid and structural studies. The chimeric alleles constructed in this study may be very useful in addressing these questions. Where HET-c is localized in the cell during vegetative growth and whether HET-c is localized in a different cellular compartment during vegetative incompatibility need to be investigated. Truncation of HET-c at the N-terminus and the association with its dual functions were proposed from the data presented in immunoprecipitation studies. These hypotheses could be examined by protein sequencing and localization studies. The HET-c:HA and HET-c:GFP fusion constructs may be useful for these studies. How vegetative incompatibility was triggered by HET-c heteromeric complex also needs to be addressed through genetic approaches, such as mutational
analysis for identifying suppressors and yeast two-hybrid analysis for identifying unknown molecules that may interact with HET-c.

In phylogenetic analyses, at least seven het-c allelic lineages were resolved among 39 naturally occurring het-c alleles (Chapter 2). However, only four types of allelic specificity, het-c\textsuperscript{OR} type, het-c\textsuperscript{PA} type, het-c\textsuperscript{GR} type and p26 (Ndi5923) were identified by the transformation assay (Chapter 3). Do the allelic lineages in phylogenetic analyses reflect functional het-c allelic types? If the specificity domain of HET-c regulates the equilibrium of the HET-c heteromeric complex formation, a heteromeric complex may be formed between HET-c proteins that were detected as the same specificity type but which were assigned to different allelic lineages. The transformation assay in this study may not be sensitive enough to identify the phenotypes resulting from such a heteromeric complex formation. A more discriminating assay, such as the yeast two-hybrid system is needed.

Morphological and biochemical evidence indicated that the vegetative incompatibility response had features in common with programmed cell death in animal and plant systems (Jacobson et al., 1998; Marek et al., 1999). The critical question that needs to be addressed towards understanding the cellular and biochemical mechanisms of vegetative incompatibility is: does vegetative incompatibility resemble the biochemical pathway as programmed cell death described in higher organisms, such as in \textit{C. elegans} or mammalian cells? If not, what are the participants in the pathway? Since \textit{N. crassa} is a model organism for genetic studies, using genetic approaches to dissect the biochemical pathway of vegetative incompatibility might be very favorable.
6. Bibliography


Felsenstein, J. 1993. PHYLIP 3.5. Department of Genetics, University of Washington. WA.


Kumar, S., K. Tamura and M. Nei, 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1. 01. The Pennsylvania State University, University Park, PA.


gene (SLG) and the S-locus related gene (SLR1) in *Raphanus sativus* L. and self-incompatible


Sansome, E. R. 1946. Heterokaryosis, mating-type factors, and sexual reproduction in

Satta, Y., W. E. Mayer and J. Klein. 1996a. Evolutionary relationships of HLA-DRB genes

Satta, Y., W. E. Mayer and J. Klein. 1996b. HLA-DRB intron 1 sequences: implications for

anserina* vegetative incompatibility locus *het-c*, whose product resembles a glycolipid
transfer protein, drastically impairs ascospore production. *Proc. Natl. Acad. Sci. USA.* 91:
5927-5931.

in the fungus *Podospora anserina* encodes a protein with a GTP-binding motif and Gß

Saupe, S., B. Turcq and J. Begueret. 1995b. Sequence diversity and unusual variability at the
Genet.* 27: 466-471.

netura of mutations in the *mt A-I* gene of the *Neurospora crassa* A idiomorph and their


### 7. Appendices

#### 7.1 DNA sequences alignment for phylogenetic analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 3</th>
<th>Sequence 4</th>
<th>Sequence 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS2740</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATTCGTTTCA GCTTGGAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB1903</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB7140</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS2741</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC4481</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1824</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI3721</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8239</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8241</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA963</td>
<td>CAAGGCTGGA GCACCACAAG TGCCGGCTAT ATTCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2739</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND3139</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND6793</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC4481</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1824</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI3721</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8239</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8241</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA963</td>
<td>CAAGGCTGGA GCACCACAAG TGCCGGCTAT ATTCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2739</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND3139</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND6793</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC4481</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1824</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI3721</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8239</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8241</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA963</td>
<td>CAAGGCTGGA GCACCACAAG TGCCGGCTAT ATTCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2739</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND3139</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND6793</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC4481</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1824</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI3721</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8239</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8241</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA963</td>
<td>CAAGGCTGGA GCACCACAAG TGCCGGCTAT ATTCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2739</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND3139</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND6793</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC4481</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1824</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI3721</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8239</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSP8241</td>
<td>CAAGGCTGGA ACACCACAAG TGCCGGCTAT ATCCGTTTCA GCTTGCAGCG TTGCATCCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.2 DNA sequences of artificial mutant \textit{het-c} alleles \textit{pd1}, \textit{pd2}, \textit{pd3}, \textit{pol} and \textit{del3}

\text{>pd1}

3' AATCGATCCTCAAAAAATCGGTTCAAGGCGCCAGCGGCTTACTTTCTTCTACGGCTGCTGAAAGCTGTCTATTCTCGAGTTTGAGTT
TAGTGTCGATGCCAATGCGGGGAAAGACTGACTCTTCTTCCATATCGATTAGGACCAATTCGCAGTAGTTACTGTGGG
CAGNAAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TAATCGGAGCCTAATTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

\text{>pd2}

3' AATCGATCCTCCTCAAAAAATCGGTTCAAGGCGCCAGCGGCTTACTTTCTTCTTCTACGGCTGCTGAAAGCTGTCTATTCTCGAGTTTGAGTT
TAGTGTCGATGCCAATGCGGGGAAAGACTGACTCTTCTTCCATATCGATTAGGACCAATTCGCAGTAGTTACTGTGGG
CAGNAAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TAATCGGAGCCTAATTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

\text{>pd3}

3' AATCGATCCTCCTCAAAAAATCGGTTCAAGGCGCCAGCGGCTTACTTTCTTCTTCTACGGCTGCTGAAAGCTGTCTATTCTCGAGTTTGAGTT
TAGTGTCGATGCCAATGCGGGGAAAGACTGACTCTTCTTCCATATCGATTAGGACCAATTCGCAGTAGTTACTGTGGG
CAGNAAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TAATCGGAGCCTAATTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

\text{>del3}

3' AATCGACCTCCTCAAAAAATCGGTTCAAGGCGCCAGCGGCTTACTTTCTTCTTCTACGGCTGCTGAAAGCTGTCTATTCTCGAGTTTGAGTT
TAGTGTCGATGCCAATGCGGGGAAAGACTGACTCTTCTTCCATATCGATTAGGACCAATTCGCAGTAGTTACTGTGGG
CAGNAAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TAATCGGAGCCTAATTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

\text{>pol}

3' AATCGATCCTCCTCAAAAAATCGGTTCAAGGCGCCAGCGGCTTACTTTCTTCTTCTACGGCTGCTGAAAGCTGTCTATTCTCGAGTTTGAGTT
TAGTGTCGATGCCAATGCGGGGAAAGACTGACTCTTCTTCCATATCGATTAGGACCAATTCGCAGTAGTTACTGTGGG
CAGNAAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TAATCGGAGCCTAATTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

7.3 Sequences of \textit{het-c}^{OR}:GFP fusion genes

\text{>ORSmGFP}

5' CCCACGCCATGCCTCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCC
TAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT

\text{>pol}

5' CCCACGCCATGCCTCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCCAGCGCC
TAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT