Allelic Specificity at the \textit{het-c} Heterokaryon Incompatibility Locus of \textit{Neurospora crassa} Is Determined by a Highly Variable Domain

Sven J. Saupe and N. Louise Glass

Botany Department and The Biotechnology Laboratory, The University of British Columbia, Vancouver, BC V6T 1W5, Canada

Manuscript received January 3, 1997
Accepted for publication April 21, 1997

ABSTRACT

In filamentous fungi, the ability to form a productive heterokaryon with a genetically dissimilar individual is controlled by specific loci termed \textit{het} loci. Only strains homozygous for all \textit{het} loci can establish a heterokaryon. In \textit{Neurospora crassa}, 11 loci, including the mating-type locus, regulate the capacity to form heterokaryons. An allele of the \textit{het-c} locus (\textit{het-c}^\textit{OR}) of \textit{N. crassa} has been previously characterized and encodes a nonessential 966 amino acid glycine-rich protein. Herein, we describe the genetic and molecular characterization of two \textit{het-c} alleles, \textit{het-c}^\textit{PM} and \textit{het-c}^\textit{EM}, that have a different specificity from that of \textit{het-c}^\textit{OR}, showing that vegetative incompatibility is mediated by multiple alleles at \textit{het-c}. By constructing chimeric alleles, we show that \textit{het-c} specificity is determined by a highly variable domain of 34–48 amino acids in length. In this regard, \textit{het-c} is similar to loci that regulate recognition in other species, such as the (S) self-incompatibility locus in plants, the sexual compatibility locus in basidiomycetes and the major histocompatibility complex (MHC) genes in vertebrates.
but strains containing an inactivated \textit{het-c}^{GR} allele display dual compatibility with both \textit{het-c}^{GR} and \textit{het-c}^{PA} strains (SAUPE et al. 1996).

In this article, we analyzed the molecular basis of allelic specificity and sequence variability at the \textit{het-c} locus. By both genetic and molecular analyses, we demonstrated that specificity is mediated by multiple alleles at \textit{het-c}. We have isolated two novel \textit{het-c} alleles, \textit{het-c}^{PM} and \textit{het-c}^{GR}, which are distinct from \textit{het-c}^{GR} and distinct from each other. By the construction of chimeric \textit{het-c} alleles, we showed that a 34–48 amino acid variable region, which is dissimilar in \textit{HET-C}^{GR}, \textit{HET-c}^{PA} and \textit{HET-c}^{GR}, controls allelic specificity.

**MATERIALS AND METHODS**

**Strains, media and nomenclature:** We introduced a nomenclature change to designate the different \textit{het-c} alleles. Historically, the two known \textit{het-c} alleles were designated \textit{het-c} and \textit{het-C}. The multiallelic nature of the \textit{het-c} locus described in this study requires a more detailed nomenclature; \textit{het-c} alleles will be designated by a two-letter acronym derived from the name of the wild-type isolate from which they originated. We designated \textit{het-c}, \textit{het-c}^{GR} (for the Oak Ridge background) and \textit{het-c}, \textit{het-c}^{PM} (for Emerson background). The \textit{het-c} base symbol now designates the locus rather than the historical \textit{het-c} allele.

A list of the \textit{N. crassa} strains used in this study is given in Table 1, together with their origin and relevant genotype. The C9 series of strains are \textit{het-c} at \textit{Het-c}, but are Oak Ridge at all other \textit{het} loci. The C2 series are either \textit{het-c}^{PA}, \textit{het-c}^{GR} or both, but are Oak Ridge at all other \textit{het} loci (SMITH et al. 1996). Vogel’s (Vogel 1964) and Westergaard’s (Westergaard and Mitchell 1947) synthetic media were used for vegetative cultures and crosses, respectively.

Crosses with the translocation strains were performed using the normal sequence strain as the female. Ascospores were heat shocked and viable progeny were allowed to germinate on Broekman and de Serres (Brockman and de Serres 1963) medium for 2 days. Inhibited progeny were scored directly on the germination plate. Counts were on a total of ~90 ascospores. For each cross, a number of inhibited progeny were then transferred to Vogel’s media plates and incubated at 30° for 2–3 days to confirm the inhibited phenotype.

**Isolation of the \textit{het-c}^{PA} and \textit{het-c}^{GR} alleles:** The \textit{het-c}^{PA} allele was isolated from the c9-2 strain (Table 1). This \textit{het-c} allele originates from the FGSC 2190 Panama strain, which is derived from FGSC 1131 Panama A wild-type isolate (PERKINS 1975; SMITH et al. 1996). The entire \textit{het-c}^{PA} allele is contained in a 4-kilobase (kb) EcoRI-XhoI fragment (SAUPE et al. 1996). A c9-2 size-selected subgenomic library was constructed by digesting 30 μg of c9-2 genomic DNA with EcoRI and XhoI. Following agarose gel electrophoresis, fragments ranging from 3.5 to 4.5 kb in size were recovered and ligated to the KSII bluscript vector (Stratagene, CA). Colony hybridization was performed on ~3000 bacterial clones using a 3.9-kb EcoRI-Sal fragment from \textit{het-c}^{GR} (SAUPE et al. 1996) as a probe.

The \textit{het-c}^{GR} allele was isolated as two overlapping PCR products. Primers were 5′(7)CTCAGATTTTGGATGATTC(26)3′ and 5′(1424)CTCACCAACACGGAGTG(1441)3′ for the 5′ end of the gene and 5′(598)GGAGACATGGCGCATCTG(615)3′ and 5′(552)GGAGACATGGCGCATCTG(3098)3′ for the 3′ end (numbers in parentheses give the position of the 5′ and 3′ base of each primer for the \textit{het-c}^{GR} sequence; SAUPE et al. 1996). The two fragments were cloned into the pCRII vector using the TA cloning kit (In Vitrogen, San Diego, CA). The \textit{het-c}^{GR} allele was then reconstructed using a unique SalI site located in the overlap region between the two fragments.

**Transformation assay for vegetative incompatibility:** \textit{N. crassa} spheroplasts were prepared according to SCHWEIZER et al. (1981). The majority of transformation experiments used c9-2 \textit{het-c}^{PA} and c2(2)9 \textit{het-c}^{GR} as recipient strains (Table 1). For all transformation experiments, \textit{het-c} alleles were cloned into a pCB1004 vector carrying the hyg 

**Construction of chimeric alleles:** For construction of \textit{het-c} chimeric alleles, both \textit{het-c}^{GR} and \textit{het-c}^{PA} were subcloned as XhoI-EcoRI fragments in the pCRII vector (In Vitrogen, San Diego, CA) to make the Std, SalI and EcoRV sites in the \textit{het-c} sequences unique (EcoRV is unique only in \textit{het-c}^{GR}). The Std-SalI and EcoRV-SalI fragments were then exchanged between the different alleles. Chimeric alleles were identified by restriction digests using a XhoI restriction site difference located within the exchanged fragment. The chimeric alleles were then re-cloned into the pCB1004 vector (CARROLL et al. 1994) to test activity in transformation assays.

The \textit{het-c}^{GR}(1–3000) allele was constructed by inserting a XhoI-EcoRV fragment encompassing 455 base pairs (bp) of the \textit{het-c}^{PA} promoter and the amino terminal 49 codons of \textit{het-c}^{PA} into the conserved EcoRV site of \textit{het-c}^{GR}.

**PCR and DNA sequence determination:** The DNA sequence of \textit{het-c}^{PA} and \textit{het-c}^{GR} was obtained using a series of overlapping subclones and by using \textit{het-specific} primers. DNA sequence was determined using the ABI automated sequencing procedure (Mississauga, Ontario, Canada) at the NAPS unit, Biotechnology Laboratory, University of British Columbia. Genomic DNA for PCR was obtained as described by IRELAND et al. (1995). For amplification of the region of specificity in FGSC 1495, FGSC 430, FGSC 1455 and FGSC 1824, primers were 5′(598)GGAGACATGGCGCATCTG(615)3′ and 5′(1424)CTCACCAACACGGAGTG(1441)3′ (numbers in parentheses give the position of the 5′ and 3′ base of each primer for the \textit{het-c}^{PM} sequence; SAUPE et al. 1996). DNA sequence determination was performed directly on gel-purified PCR amplification products.

**RESULTS**

**Genetic evidence for multiple allelism at \textit{het-c}** The \textit{het-c}^{GR} (PA for Panama) was used as an antagonistic allele in the isolation and the molecular characterization of \textit{het-c}^{GR} (\textit{het-c}) (SAUPE et al. 1996). Both heterokaryon tests and transformation studies indicate that \textit{het-c}^{PA} is incompatible with \textit{het-c}^{GR} (SAUPE et al. 1996).

In earlier genetic studies (PERKINS 1975), the specificity of \textit{het-c}^{PA} with respect to \textit{het-c}^{PM} (\textit{het-c}) and \textit{het-c}^{GR} translocation testers could not be determined because of the linkage of a second vegetative incompatibility locus \textit{het-f}, which confounded genetic analyses. The availability of near-isogenic strains that differed at only \textit{het-c} (SMITH et al. 1996) allowed us to examine the \textit{het-c} specificity of an inbred \textit{het-c}^{PA} strain C9-2 (Table 1). As expected, approximately one third of the viable progeny displayed a \textit{het-c} incompatible phenotype in crosses between the
TABLE 1  
Neurospora crassa strains

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Genotype</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>c9-2</td>
<td>thr-2 het-cPA het-6WM a</td>
<td>M. L. SMITH</td>
</tr>
<tr>
<td>c(2)-2-1</td>
<td>thr-2 het-cRO het-6PA a</td>
<td>M. L. SMITH</td>
</tr>
<tr>
<td>c(2)-2-5</td>
<td>thr-2 het-cRO het-6PA a</td>
<td>M. L. SMITH</td>
</tr>
<tr>
<td>c(2)-2-9</td>
<td>thr-2 het-cRO het-6PA a</td>
<td>M. L. SMITH</td>
</tr>
<tr>
<td>FGSC 847</td>
<td>A</td>
<td>Lein-7</td>
</tr>
<tr>
<td>FGSC 1693</td>
<td>a</td>
<td>Lein-8</td>
</tr>
<tr>
<td>FGSC 1924</td>
<td>A</td>
<td>Lahore-1</td>
</tr>
<tr>
<td>FGSC 450</td>
<td>A</td>
<td>Adipodoume</td>
</tr>
<tr>
<td>FGSC 4711</td>
<td>A</td>
<td>Haut Diquini</td>
</tr>
<tr>
<td>FGSC 1945</td>
<td>a</td>
<td>Groveland 1-c</td>
</tr>
<tr>
<td>FGSC 967</td>
<td>a</td>
<td>Liberia</td>
</tr>
<tr>
<td>FGSC 1455</td>
<td>het-cEM het-D; het-2; ini A</td>
<td>J. F. WILSON</td>
</tr>
<tr>
<td>FGSC 3879</td>
<td>T(III &gt; VR)NM149 het-cOR A</td>
<td>D. D. PERKINS</td>
</tr>
<tr>
<td>FGSC 1483</td>
<td>T(III &gt; VR)NM149 het-cEM A</td>
<td>D. D. PERKINS</td>
</tr>
<tr>
<td>FGSC 2193</td>
<td>T(III &gt; VR)NM149 het-cGR A</td>
<td>D. D. PERKINS</td>
</tr>
<tr>
<td>FGSC 2191</td>
<td>T(III &gt; VR)NM149 het-cGR A</td>
<td>D. D. PERKINS</td>
</tr>
</tbody>
</table>

C9-2 (het-cPA) and T(III > VR)NM149 het-cOR translocation tester strains. Similarly, in crosses between C9-2 (het-cPA) and the T(III > VR)NM149 het-cEM translocation tester, approximately one third of the viable progeny also displayed a het-c incompatibility phenotype. These results indicated that het-cPA represented a different allelic specificity from both het-cOR and het-cEM. In this regard, the het-cPA allele genetically resembled the het-cOR allele (PERKINS 1975), which was also determined to be distinct from both het-cOR and het-cEM. To determine if het-cPA had a similar het-c specificity to het-cEM or represented a fourth allele at het-c, the C9-2 strain was crossed to the T(III > VR)NM149 het-cOR translocation strain. Progeny that exhibited a het-c incompatibility phenotype were not observed, indicating that het-cPA and het-cEM had an identical het-c specificity.

The allelic specificity at het-c of Groveland 1-c isolate (FGSC 1445) could not be determined previously because of differences at het-6 between Groveland 1-c and the T(III > VR)NM149 het-cOR and het-cEM testers (which are het-cEM) (PERKINS 1975). The availability of the C(2) testers (Table 1) that are het-cPA (SMITH et al. 1996) allowed us to determine the het-c specificity of Groveland 1-c isolate. In crosses between T(III > VR)NM149 het-cOR and the C(2)-2-1 (het-cOR het-cPA) and C(2)-2-5 (het-cPA het-cOR) strains, approximately one third of the viable progeny displayed a typical het-c incompatibility phenotype. These results indicated that Groveland 1-c was distinct in het-c specificity (het-cGR) from both het-cPA and het-cOR strains.

Isolation and characterization of the het-cPA and het-cGR alleles: The het-cPA allele has a high degree of sequence similarity to het-cGR based on DNA-DNA hybridization studies (SAUPE et al. 1996). To isolate the het-cPA allele, a subgenomic library from the c9-2 het-cPA strain was constructed (see MATERIALS AND METHODS). The sequence of the 4 kilobase pair (kbp) insert of one clone was determined. The het-cGR allele was isolated by PCR, using primers chosen in regions that were conserved between het-cGR and het-cPA. The sequence of the entire 3.8-kbp het-cGR product was determined.

The putative promoter region of het-cGR differed from both het-cOR and het-cPA by a 3-bp deletion in a repeated partial palindrome (5' CTA/cTAG 3') (Figure 1). This sequence was located 15 bp upstream of the putative transcription start site as defined by the TCATCANC consensus (BRUCHETZI et al. 1993). This region was determined to be important for het-cOR expression as a Thr5 insertion immediately upstream of the palindrome abolishes het-cOR activity (SAUPE et al. 1996). The effect of promoter differences between het-cGR and het-cPA was investigated by transformation experiments (see below).

Sequences defined as the 5' donor, 3' acceptor and lariat formation sites for the het-cOR introns (SAUPE et al. 1996) were conserved in the het-cPA and het-cEM alleles (data not shown). The het-cPA and het-cGR alleles had an open reading frame (ORF) of 974 and 963 amino acids, respectively (Figure 2). An amino acid comparison between HET-cOR, HET-cPA and HET-cGR showed 86% identity between the three polypeptides. However, the degree of conservation between HET-cOR, HET-cPA and HET-cGR varied along the amino acid sequence. Two regions that were highly similar between the three polypeptides were observed. The first region encompassed the 148 N-terminal amino acids, which were 99% identical between the three polypeptides; only two amino acid substitutions were present in HET-cPA (positions 2 and 5) (Figure 2). These substitutions would not affect the function of the putative signal sequence (positions 1–31, Figure 2; GIERASCH 1989; SAUPE et al. 1996). Eleven additional nucleotide substitutions in this region were present in het-cPA and seven were present in het-cGR, six of which were identical to those in het-cPA. None of these nucleotide substitutions resulted in amino acid changes (data not shown).

A second 200 amino acid region (positions 285–486), which encompassed the putative coiled-coil domain (positions 426–458; Figure 2; SAUPE et al. 1996), was also highly conserved between HET-cOR, HET-cPA and HET-cGR (99% identity). The amino acid sequence of the coiled-coil domain was identical between HET-cOR and HET-cPA; HET-cGR had one conservative amino acid substitution. In the conserved 200 amino acid region, het-cPA had 22 nucleotide substitutions, only one of which resulted in a conservative amino acid substitution. The het-cGR sequence differed from het-cOR at 43 positions, 14 of which were identical to the het-cPA substitutions. These changes resulted in only three amino acid differences, two of which were conservative substitutions.

In contrast to conserved regions between HET-cOR, HET-cPA and HET-cGR, four regions displayed signifi-
cant variability. In the first region (region I), from amino acid positions 149–212, the identity between HET-cGR and HET-cPA was 78% (Figures 2 and 3). The het-cPA sequence differed from het-cGR at 23 positions, resulting in 11 amino acid differences, only four of which were conservative substitutions. The het-cGR sequence varied from the het-cOR sequence at 52 positions; 21 of which were at identical positions to those observed in het-cPA. The amino acid sequence of HET-cGR varied from HET-cOR at 14 positions, 10 of which were identical differences observed in HET-cPA.

A second region (II) from positions 247 to 284, showed only 27% identity between the three polypeptides (Figures 2 and 3). Within this 38 amino acid region, the HET-c polypeptides differed by 20 amino acid substitutions and three deletion (or insertions) events. The het-cPA allele had a 45-bp insertion that resulted in a 15 amino acid insertion that was absent in HET-cOR and HET-cGR. The inserted DNA sequence was unrelated to any adjacent DNA sequences. A 15-bp insertion in het-cGR, 9 bp upstream from the het-cPA-specific insertion, resulted in a five amino acid insertion that was absent in HET-cPA and HET-cGR (Figure 2). In a region located immediately upstream of the insertion in het-cGR, nine out of 15 nucleotide positions were identical to the insertion. Thus, the het-cGR-specific insertion could have been generated by a 15-bp duplication with subsequent nucleotide pair substitutions.

A third highly variable region was identified in the HET-c polypeptides (region III) from amino acid position 487–610; (Figures 2 and 5). Amino acid identity between HET-cOR and HET-cPA was 96% in this region, while identity with HET-cGR was only 68%. A comparison of the DNA sequences of the three alleles in this region suggested that at least three deletion or insertion events occurred in het-cGR with additional base pair substitutions.

The glycine-rich domain (region IV) described in HET-cGR (SAUPE et al. 1996) (amino acid positions 610–936) was also quite variable (87% identity; Figures 2 and 3). The three polypeptides differed by 32 amino acid substitutions as well as in the number of glycine-rich repeats. HET-cPA had a six amino acid insertion of a GG YGGY repeat at position 841 and a deletion of an eight amino acid GG YGGY repeat at position 936 as compared to the HET-cOR (Figure 2). HET-cGR also had an insertion at position 841 and a deletion near position 936, although the insertion and deletion were of a different size from that of HET-cPA. This type of allelic polymorphism involving variability in the number of repeats is common in genes containing repetitive regions (JARMAN and WELLS 1989; SAUPE et al. 1995a) and is believed to result from unequal crossing over in the repeated regions.

**Transformation assays for incompatibility with the het-cPA and het-cGR alleles:** The het-cPA allele was subcloned into the hygromycin resistance vector pCB1004 (CARROLL et al. 1994) for transformation experiments. When the het-cPA construct was introduced into c(2)-2-9 het-cOR (Table 1) spheroplasts, >80% of the transformants showed a strong growth inhibition and an altered colony morphology (Figure 4). In contrast, the introduction of the same construct into the c9-2 het-cPA spheroplasts did not affect growth or morphology of the transformants. The phenotype of a c(2)-2-9 het-cOR strain transformed with the het-cPA allele was slightly different from the phenotype of a c9-2 het-cPA strain transformed with the het-cGR allele. The slow-growing mycelium of the c9-2 het-cPA (het-cGR) transformants appeared compacted into curled macroscopic branches (Figure 4). In the c(2)-2-9 het-cGR (het-cPA) transformants, the slow-growing mycelium was straighter but more sparse than the het-cPA (het-cGR) transformants and radial growth was more prominent.

The introduction of het-cGR into het-cPA spheroplasts gave transformants that displayed the phenotype of het-cPA (het-cGR) transformants, although somewhat less severe (Figure 4). However, the introduction of het-cGR into het-cOR spheroplasts gave transformants that exhibited only a slight growth inhibition and limited condensation. To determine if the deletion found in the het-cGR promoter sequence affected het-c activity, the promoter and the first 49 codons of het-cGR were replaced by the corresponding sequences from het-cPA (Figure 1; see MATERIALS AND METHODS). The construct HET-cGR(PA149-40) differed from HET-cGR by the two conservative amino acid substitutions in the signal peptide (positions 2 and 5) in addition to the promoter sequences. The assumption was made that these amino acid changes are nonrelevant to specificity based on the data on location of the specificity region presented in the following section. The introduction of the het-cGR(PA149-40) construct into het-cPA spheroplasts gave inhibited transformants similar in phenotype to het-cPA (het-cGR) transformants.

**FIGURE 1.—Comparison of het-c sequences in promoter regions.** Allele designation is given on the left. Numbering corresponds to the het-cPA sequence. Position of the Tn5 insertion point is given (SAUPE et al. 1996). The palindromic repeat appears in boldface and the putative transcriptional start site for het-cPA is underlined. * indicates conserved positions.
FIGURE 2.—Amino acid sequence comparison of the \( \text{het-c}^{GR} \), \( \text{het-f}^{A} \) and \( \text{het-c}^{CR} \) translated products. Allele designation is given on the left. Sequences are compared to the \( \text{het-c}^{GR} \) product. Periods indicate identity and dashes indicate gaps. Numbering corresponds to the \( \text{het-c}^{GR} \) product. Sequence alignment was generated by using the pattern-induced multisequence alignment (PIMA) algorithm (Smith and Smith 1992) with manual adjustments.
signal peptide | coiled-coil domain | glycine-rich domain
--- | --- | ---
I | 149-212 | 487-610 |
II | 247-284 | 708 |
III | 966 | 1149-212
IV | 247-284 | 487-610 |
966 aa

**Figure 3.** Structure and specificity of chimeric *het-c* alleles. A schematic representation of the *het-c* product is given and the identified sequence features are highlighted. The results of the introduction of the chimeric alleles into *het-c*PA and *het-c*OR recipients is given on the right (+, incompatible interaction; −, compatible interaction; +/- a proportion of the transformants, 20–50%, show some morphological alterations and growth inhibition).

(Figure 4). When *het-c*GR(PA1-49) was introduced into *het-c*GR spheroplasts, the phenotype of the transformants was more severe than those obtained with *het-c*GR, however, the growth inhibition was not as severe as in the *het-c*PA[*het-c*GR(PA1-49)] transformants. These results suggested that the deletion in the partial palindrome repeat in the *het-c*GR promoter caused a reduction in *het-c*GR expression and thus attenuated the incompatibility response as compared to that of the *het-c*GR(PA1-49) construct.

The introduction of the *het-c*GR(PA1-49) construct into *het-c*PA spheroplasts produced a strong incompatibility phenotype and indicated that *het-c*GR and *het-c*PA were distinct alleles. This was consistent with the fact that *het-c*GR/*het-c*PA partial diploids were incompatible. However, the introduction of *het-c*GR(PA1-49) into *het-c*OR spheroplasts resulted in only a partial growth inhibition, whereas *het-c*OR/*het-c*GR partial diploids were severely inhibited in growth (see data from previous section). We therefore introduced the cloned *het-c*GR and *het-c*OR alleles into spheroplasts of FGSC 1945 *het-c*GR. The *het-c*GR/*het-c*GR transformants were severely inhibited in growth and displayed a typical incompatible *het-c* phenotype while the *het-c*GR/*het-c*GR transformants grew well and were similar in phenotype to transformants containing only the vector (Figure 5). This experiment demonstrated that *het-c*GR was also clearly distinct in specificity from *het-c*OR. Thus, the transformation assays confirmed that *het-c*OR, *het-c*GR and *het-c*PA have different *het-c* specificities.

**Specificity at *het-c*** is mediated by a highly variable

---

**Figure 4.** Transformation of c9-2 (*het-c*PA) and c(2)-2-9 (*het-c*OR) recipients with the cloned *het-c*PA, *het-c*OR, *het-c*GR and *het-c*GR(PA1-49) alleles. Transformants were transferred to plates containing hygromycin and incubated at 30° for 3 days. Designation of the recipient strain and the introduced allele are given.
domain: To dissect the region that determined het-e specificity, we constructed chimeric alleles between the het-eGR and het-ePA sequences and tested them for activity in a transformation assay. Previous studies had shown that the C-terminal portion of the glycine-rich domain (amino acid positions 729–966) was dispensable for het-eGR activity (Saupe et al. 1996). Therefore, it was unlikely that amino acid differences between HET-eGR and HET-ePA downstream of amino acid position 729 were critical for determining het-e specificity.

We chose to make chimeric constructs that included regions I and II. We exchanged a 801-bp EcoRV-SalI fragment from het-eGR that included both region I and II variable regions by the corresponding sequence from the het-ePA allele (Figures 2 and 3). The het-eORPA50–294 chimeric allele was introduced into both het-ePA (C9-2) and het-eGR (c(2)-2-1) spheroplasts (Table 1). The chimeric construct gave incompatible transformants when introduced into het-eGR but not when introduced into het-ePA spheroplasts (Figure 3). This result indicated that differences within the region from positions 50 to 294 were sufficient to confer het-ePA specificity.

To further dissect the specificity region, two additional chimeric alleles were constructed by exchanging a 216-bp StuI-SaII fragment (encompassing only region II) between the two alleles (Figure 3). Region II is the most divergent between HET-eGR and HET-ePA. The corresponding chimeric alleles het-eORPA223–294 and het-ePA00223–294 were introduced into het-eGR and het-ePA spheroplasts. The het-eORPA223–294 chimeric allele conferred het-eGR activity whereas the het-ePA00223–294 chimera conferred het-eGR activity (Figures 3 and 6). The inhibited growth phenotype of transformants containing the chimeras reflected the origin of the specificity region, i.e., the phenotype of the het-ePA[het-ePA00223–294] transformants resembled the morphology of het-ePA[het-eGR] transformants while the sparse growth of the het-eOR[het-eORPA223–294] transformants was similar to the het-eOR[het-ePA] transformants. Thus, the simple exchange of a 216-bp StuI-SaII fragment (region II) conferred het-e specific allelic specificity in the two reciprocal chimeras.

To examine if the region of het-eGR specificity was also determined by region II, we constructed a chimeric allele between het-eGR and het-ePA. We exchanged the 216-bp StuI-SaII fragment of het-ePA with the corresponding fragment from the het-eGR allele (Figure 3). The corresponding chimeric construct het-ePA[het-ePA00223–294] was introduced into het-eGR and het-eOR spheroplasts. The het-ePA, [het-ePA[het-ePA00223–294]] transformants were clearly incompatible and similar in phenotype to het-ePA[het-ePA00223–294] transformants (Figures 3 and 6). The het-eGR[het-ePA00223–294] transformants were also inhibited and similar in phenotype to those obtained with the het-ePA[het-ePA00223–294] construct. These results indicated that the region from amino acid positions 223 to 294 was sufficient to determine het-eGR specificity. Transformation assays with chimeric alleles showed that het-e specificity is determined by a restricted and highly variable region (from positions 247 to 284 in HET-eGR).

Correlation between sequence variability and allelic specificity in other wild-type isolates: To determine if a correlation existed between sequence variability observed in region II and het-e specificity as determined by genetic analyses, we amplified this region by PCR from a number of N. crassa isolates that had been previously characterized genetically for specificity at het-c (Perkins 1975; D. D. Perkins, personal communication and this study). These included the Adiopodoume (FGSC 430), Groveland (FGSC 1445), cDE inl (FGSC 1455), Liberia (FGSC 967), Lein-7 (FGSC 1693), Lahore-1 (FGSC 1824) and Haut Diquini (FGSC 4711) strains (Table 1). Figure 7 shows the amino acid sequence from amino acid positions 222 to 294 for these isolates together with their het-e specificity as determined genetically in crosses with the het-eGR and het-eOR translocation testers.

The DNA and amino sequence of the specificity region in Adiopodoume FGSC 430 was identical to that of C9-2 het-ePA. Genetic analyses with partial diploids indicated that het-ePA and het-eOR had an identical het-e specificity (see previous section). The het-ePA and het-eOR alleles were also determined to be distinct from both het-eGR and het-eGM (this study; Perkins 1975). A transformation experiment was performed in which the het-ePA and het-eOR alleles were introduced into FGSC 430 het-eGR spheroplasts. Transformants displaying het-e incompatibility were not observed among the het-eOR[het-eGM] transformants, but the het-eOR[het-eGM] transformants displayed typical het-e incompatibility (data not shown). Therefore, genetic analysis, transformation assays and sequence analysis all indicated that het-eOR had the same allelic specificity as het-ePA.

The specificity region from three isolates that were genetically characterized as being het-eGR were similar in variable region motif to that of HET-eGR (Figure 7). All three isolates were missing the 15 amino acid insertion of HET-ePA, but retained the five amino acid insertion that was characteristic of HET-eGR. The FGSC 1693 (Lein-8) varied from HET-eGR at only two amino acid positions in the het-e specificity region. However, HET-
The third type of motif in the variable domain was exhibited by three strains, FGSC 1945 (Groveland I-c; het-c OR), FGSC 1455 (het-c EM) and FGSC 967 Liberia (het-c EM). HET-c OR, HET-c EM and HET-c LB were missing the 15 amino acid HET-c PA-specific insertion but were also lacking the five amino acid HET-c OR-specific insertion (Figure 7). HET-c EM and HET-c LB were identical in amino acid sequence in the specificity region and differed from HET-c OR by only one amino acid substitution. The FGSC 1455 (het-c EM) strain has been defined as a het-c tester and was the original antagonistic allele as described by GARNJOBST and WILSON (1956). In previous studies (PERKINS 1975), it was determined that the het-c EM strain and the Liberia isolate (FGSC 967) have identical het-c specificity. The het-c OR allele genetically resembles het-c EM as it is incompatible both with het-c OR and het-c PA. It is likely that het-c OR and het-c EM have the same het-c specificity.

HET-c OR, HET-c EM and HET-c OR contained a region that was almost identical in amino acid sequence to the polymorphic region identified in HET-c LA (positions 247–268) (Figure 7). The presence of the polymorphic stretch in HET-c LA may have resulted from a crossover between a het-c EM type allele and a het-c OR type allele. However, as mentioned previously, this region cannot be required for conferring het-c specificity. The results of this analysis indicated that sequence variability within region II and het-c specificity were correlated, supporting the transformation data indicating that this region controls het-c allelic specificity.

**DISCUSSION**

The het-c locus is one of 11 loci known to control heterokaryon formation in *N. crassa* (PERKINS 1988). In this study, we determined that specificity at het-c is mediated by three allelic types, het-c OR, het-c PA and het-c EM, and identified a variable region of the het-c encoded polypeptide that is responsible for allelic specificity.

**Multiple alleles at het-c** Sequence analysis, translocata-

![Image](het-c OR(223-294) het-c PA(222-294) het-c PA(222-294))

**Figure 6.**—The introduction of chimeric alleles into c9-2 (het-c OR) and c(2)-2-9 (het-c EM) recipients. Individual transformants were transferred to plates containing hygromycin and incubated at 30' for 3 days. Designation of the recipient strain and the introduced allele are given.

**Figure 7.**—Amino acid sequence comparison of het-c products from different *N. crassa* isolates. The strain from which the sequence was obtained is given on the left. The het-c specificity as determined by translocation crosses is given on the right (PERKINS 1975; D. D. PERKINS, personal communication and this study). Numbering is for the het-c OR product. Dashes indicate deletions and asterisks indicate conserved positions. Sequence alignment was done using the PIMA (SMITH AND SMITH 1992) algorithm with manual adjustments.
tion crosses and transformation assays indicate that het-cGR, het-cRA and het-cGR are distinct alleles. Our results suggest that het-cGR has the same specificity as het-cRM, one of the two originally defined alleles (Garnjobst and Wilson 1956). The het-cPM allele is identical in the specificity region to het-cRA, the suspected third allele in previous genetic studies (Perkins 1975; Howlett et al. 1993). Sequencing and PCR analysis of the het-c specificity region from a number of wild-type isolates revealed that these sequences fell into three distinct classes as defined by the het-cRM, het-cRA and het-cGR-type (this study and unpublished results). Thus, it appears that het-c specificity is mediated by three allelic types that have characteristic sequence motifs within the specificity region.

It is possible that multiple allelism is not restricted to het-c but may be a common feature of several het loci. Based on genetic analyses, multiple allelism has also been suggested for a het locus in Aspergillus nidulans (Dales et al. 1993) and at het-8 in N. crassa (Howlett et al. 1993). Multiple allelism has also been shown for het loci involved in nonallelic incompatibility in P. anserina (Saue et al. 1995a, b). If the function of het loci is to limit heterokaryosis, multiple allelism might improve the efficiency of incompatibility systems by increasing the number of vegetative compatibility groups and thus decreasing the likelihood of forming a heterokaryon with a genetically different individual. In other loci that regulate self/nonself recognition, such as the mating type loci of basidiomycetes (Kronstad and Leong 1990), S (self-incompatibility) locus in flowering plants (Thompson and Kirch 1992; Kothé 1996) and the histocompatibility complex (MHC) genes in vertebrates (Ayala et al. 1994), large allelic series mediate recognition or incompatibility. The results presented in this study suggest that only three allelic specificities occur at het-c. However, the high diversity of heterokaryon incompatibility groups observed in N. crassa is achieved by the presence of numerous unlinked het loci. In a survey of N. crassa isolates from three separate locations in Louisiana, all were found to be heterokaryon incompatible (Milner 1976). Assuming that all 11 het loci in N. crassa function to restrict heterokaryosis, the probable number of compatibility groups in a population, even in the absence of multiple allelism, is enormous.

Our results from this study indicated that genetic background can modulate the incompatibility response mediated by het-c. The phenotype of the het-cGR/het-cGR transfectants was less severe than that of het-cRM/het-cRM duplication progeny. However, the phenotype of the het-cRA/het-cRA transfectants was similar to that of het-cRM/het-cRM duplication progeny. These results indicate that genetic background significantly affects the interaction of het-cGR with distinct nonallelic gene in Groveland 1-c. het genes involved simultaneously in allelic and nonallelic interaction have been described in P. anserina (Begueret et al. 1994).

**Specificity region:** The construction of chimeric alleles showed that the difference in a region from amino acid positions 247 to 284 was sufficient to determine het-c specificity. The polymorphism in this region resulted from at least two independent insertion (or deletion) events; one generating the five amino acid het-cRM specific region and the second generating the 15 amino acid het-cRA-specific region. The het-cGR-specific insertion is not required to trigger incompatibility with het-cRA, because het-cGR (which is missing the insertion) is incompatible with het-cRA. Rather, it is the absence of the 15 amino acid het-cRA-specific region that appears to determine incompatibility with het-cRA. Although we cannot rule out that specificity is determined by one or a few discrete positions within the defined region, the high degree of polymorphism makes it more likely that specificity is determined by the global structure of the region. This hypothesis is also supported by the fact that the HET-cLA and HET-cGR differ at numerous positions, although they have the same specificity when tested as partial diploids. It is possible that specificity is regulated by controlling the spacing of critical conserved regions on either side of the polymorphic stretch. A database search of the HET-cRA specific insertion did not reveal any significant similarity with known protein domains. However, the het-cRM-specific insertion contains a N-glycosylation site (NDT) and it may be that the het-c products differ in glycosylation patterns.

The most variable region of het-c was found to control specificity. Localization of specificity determinants in highly polymorphic regions has been described in a number of loci involved in recognition, such as the b mating type locus from the basidiomycetes Ustilago maydis (Dahl et al. 1991; Yee and Kronstad 1993) or Coprinus cinereus (Gieser and May 1994), the S locus in plants (Nasrallah et al. 1987; Anderson et al. 1989; Kaufmann et al. 1991; for review see Clarke and Newbigin 1993) and the MHC genes in vertebrates (Bjorkman and Parham 1990). The het-c products are glycine-rich proteins that are though to be located in the fungal cell wall. It is possible that they are organized in a higher order structure and that assembly is disrupted when distinct het-c products co-exist in the cell wall. In this hypothesis, het-c incompatibility would represent a direct consequence the co-expression of distinct HET-c proteins.

**Variable degree of polymorphism:** The identity between the three het-c polypeptides is relatively low (86%, 108 substitutions and eight insertions or deletions). Because a strain containing a het-c RIP-inactivated allele displayed no deleterious vegetative or sexual phenotype and that the severe phenotype observed when het-cGR is introduced into the het-cGR strain actually corresponds to the interaction of het-cGR with a distinct nonallelic gene in Groveland 1-c, het genes involved simultaneously in allelic and nonallelic interaction have been described in P. anserina (Begueret et al. 1994).
(Saupe et al. 1996), it is possible that this relatively high variability results from the absence of any selective pressure to maintain the protein sequence. However, two large regions of the protein show strong conservation at the amino acid level (from positions 1 to 148 and from positions 285 to 486). Moreover, single base pair deletions or insertions are frequent in noncoding regions, but all deletion and insertion events in the het-c ORF maintain the same reading frame. These observations strongly suggest that a selective pressure exists to maintain het-c.

The fact that the het-c allele comparison reveals that the specificity is determined by the most variable region suggests that positive selection is acting to create and maintain a polymorphism in this particular region of het-c. This data suggests that het-c may be required to limit heterokaryon formation in natural populations and that there is a selective advantage in avoiding heterokaryosis. The existence of a selective pressure promoting polymorphism has also been postulated to explain unusual sequence variability at another het locus, the het-c locus involved in nonallelic incompatibility in P. anserina (Saupe et al. 1995b). Positive selection leading to high variability in specificity domains appears to be a common feature in the evolution of several recognition systems including the peptide binding region of the MHC genes (Hughes and Nei 1988; Lundberg and Devitt 1992). Analyses of het-c variability in natural Neurospora populations are required to explore and expand this hypothesis of selected polymorphism in fungal recognition systems. Molecular identification of the het-c specificity region will be a valuable tool in those studies examining the evolution of self/nonself recognition.

The authors thank Katherine Tung for technical assistance in the chimeric allele construction and Jennifer Wu for help with alignment of het-c. We acknowledge that some of the translocation crosses suggesting that het-c and het-d were distinct alleles were originally done by G. A. Kuldau. We thank D. D. Perkins and R. B. Todd for their comments on this manuscript and acknowledge the continuing collaboration on vegetative incompatibility with D. D. Perkins. This work was funded by a research grant to N. L. G. from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

LITERATURE CITED


Saupe, S., B. Turcq and J. Begueret, 1993a A gene responsible
for vegetative incompatibility in the fungus *Podospora anserina* encodes a protein with a GTP-binding motif and Gg homologous domain. Gene 162: 135–139.


Communicating editor: R. H. DAVIS