Multilocus Self-Recognition Systems in Fungi as a Cause of Trans-Species Polymorphism

Christina A. Muirhead,* N. Louise Glass† and Montgomery Slatkin*,[1]

*Department of Integrative Biology, †Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3140

ABSTRACT

Trans-species polymorphism, meaning the presence of alleles in different species that are more similar to each other than they are to alleles in the same species, has been found at loci associated with vegetative incompatibility in filamentous fungi. If individuals differ at one or more of these loci (termed *het* for heterokaryon), they cannot form stable heterokaryons after vegetative fusion. At the *het-c* locus in *Neurospora crassa* and related species there is clear evidence of trans-species polymorphism: three alleles have persisted for ~30 million years. We analyze a population genetic model of multilocus vegetative incompatibility and find the conditions under which trans-species polymorphism will occur. In the model, several unlinked loci determine the vegetative compatibility group (VCG) of an individual. Individuals of different VCGs fail to form productive heterokaryons, while those of the same VCG form viable heterokaryons. However, viable heterokaryon formation between individuals of the same VCG results in a loss in fitness, presumably via transfer of infectious agents by hyphal fusion or exploitation by aggressive genotypes. The result is a form of balancing selection on all loci affecting an individual's VCG. We analyze this model by making use of a Markov chain/strong selection, weak mutation (SSWM) approximation. We find that trans-species polymorphism of the type that has been found at the *het-c* locus is expected to occur only when the appearance of new incompatibility alleles is strongly constrained, because the rate of mutation to such alleles is very low, because the number of possible incompatibility alleles at each locus is restricted, or because the number of incompatibility loci is limited.

DISTINGUISHING self from nonself is essential for sexual reproduction, for defense against pathogen invasion, and for the maintenance of individuality. Population genetics theory has been applied successfully to model the evolution of two such recognition systems, the major histocompatibility complex (MHC) of vertebrates and self-incompatibility (SI) loci in plants (WRIGHT 1939; TAKAHATA 1990). The MHC is an array of linked loci that are important in pathogen recognition (BJORKMAN and PARHAM 1990). Sexual reproduction in many plant species is mediated by gametophytic or sporophytic SI loci that elicit recognition and rejection of pollen from individuals with the same or similar genotypes (CLARKE and NEWBIGIN 1993).

Although MHC and SI loci encode different gene products, they share common evolutionary features. Numerous alleles at both kinds of loci show evidence of long-term persistence; alleles from one species are often more closely related to alleles in a closely related species than to other alleles in the same species (FIGUEROA *et al.* 1988; IOERGER *et al.* 1990), a phenomenon referred to as "trans-species polymorphism" (KLEIN 1980; ARDEN and KLEIN 1982; KLEIN *et al.* 1998). The observation of trans-species polymorphism, along with the relatively even distribution of allele frequencies found within populations, supports the idea that these loci are under some form of balancing selection, either because of overdominance in fitness or frequency-dependent selection favoring rarer alleles or genotypes. There is a well-developed population genetics theory of balancing selection at single loci that predicts the occurrence of trans-species polymorphism whenever selection is strong relative to genetic drift (TAKAHATA 1990; VEKEMANS and SLATKIN 1994). Although interactions are known to occur among MHC and SI loci, single-locus theory provides a reasonable first approximation and has led to a greater understanding of how balancing selection, genetic drift, and mutation together determine allele numbers and allele frequencies at these loci. Models of balancing selection at two closely linked loci in finite populations are more complicated but lead to predictions that are similar to those from one-locus models (KELLY and WADE 2000; SLATKIN 2000).

Trans-species polymorphism has also been detected at vegetative incompatibility loci in filamentous fungi. Nonself recognition is thought to be important during vegetative growth. As filamentous fungi grow, fusions between hyphae occur (a process called anastomosis), which yields a network of interconnected hyphae, or mycelium, that makes up the fungal individual (GLASS *et al.* 2000). Filamentous fungi also possess the remarkable

[1] *Corresponding author:* Department of Integrative Biology, University of California, Berkeley, CA 94720-3140.
E-mail: slatkin@socrates.berkeley.edu

attribute of being able to undergo hyphal fusion between different individuals to form vegetative heterokaryons that contain genetically distinct nuclei within a common cytoplasm. Heterokaryon formation has potential benefits of functional diploidy and mitotic genetic exchange ("parasexual cycle"; PONTECORVO 1956) in these haploid organisms. However, the formation of viable heterokaryons is believed to be virtually excluded in nature by the action of genetic differences at heterokaryon incompatibility (het) loci [MYLYK 1976; PANDIT and MAHESHWARI 1996; also termed vegetative incompatibility (vic) loci; LESLIE 1993]. Hyphal anastomosis between individuals that have alternative allelic specificities at any het locus usually results in compartmentation and death of the hyphal fusion cell, a phenomenon referred to as either vegetative or heterokaryon incompatibility (GLASS and KULDAU 1992; LESLIE 1993; SAUPE 2000).

Vegetative incompatibility has been described in numerous filamentous ascomycetes and basidiomycetes (GLASS and KULDAU 1992; LESLIE 1993; WORRALL 1997). Genetic studies in several ascomycete species show that there are several effectively unlinked het loci: at least 11 in Neurospora crassa (PERKINS and DAVIS 2000), 9 in Podospora anserina (SAUPE et al. 2000), 8 in Aspergillus nidulans (JINKS et al. 1966; ANWAR et al. 1993), and 7 in Cryphonectria parasitica (ANAGNOSTAKIS et al. 1986; CORTESI and MILGROOM 1998). Generally, only two or three alternative allelic specificities are found at each het locus. Two types of genetic systems have been described that regulate vegetative incompatibility, allelic and nonallelic (GLASS et al. 2000; SAUPE 2000). In allelic systems, hyphal anastomosis between individuals that contain alternative specificities at a single het locus triggers vegetative incompatibility. In nonallelic systems, an interaction between specific alleles at two different het loci mediates vegetative incompatibility. In N. crassa, C. parasitica, and A. nidulans allelic systems have been described, whereas in P. anserina both allelic and nonallelic systems have been characterized. In this article, we are concerned with allelic systems such as that found in N. crassa.

In N. crassa, 11 het loci have been genetically characterized (PERKINS and DAVIS 2000) and 3 loci have been molecularly characterized. Most loci are biallelic, although het-c and het-8 encode three alternative allelic specificities (HOWLETT et al. 1993; SAUPE and GLASS 1997). N. crassa populations are polymorphic at het loci (MYLYK 1976; DEBETS et al. 1994). One of the loci characterized molecularly, het-c, encodes three alternative allelic specificities that show trans-species polymorphisms (SAUPE and GLASS 1997; WU et al. 1998). The trans-specific lineages are defined by insertion/deletions and are thus unlikely to be the result of convergent evolution of independent mutations. Phylogenetic analysis of other markers showed that species with trans-specific het-c lineages were monophyletic (SKUPSKI et al.

1997) and that genealogies of alleles within a functional het-c class were concordant with the inferred pattern of speciation (POWELL et al. 2001), indicating that horizontal transfer is an unlikely explanation for trans-species polymorphisms at het-c. This trans-species polymorphism suggests that some form of balancing selection is operating at het-c, a suggestion further supported by the finding that individuals containing alternative specificities were equally frequent in Neurospora populations (WU et al. 1998). Genetic analysis of an N. crassa population also showed equal frequencies of molecular polymorphisms associated with the two alternative het-6 specificities (MIR-RASHED et al. 2000), although similar patterns are not always found at fungal incompatibility loci (MILGROOM and CORTESI 1999).

The role of selection in shaping fungal incompatibility systems is debated. The principal question is whether the incompatibility phenotype is itself favored by selection or whether it is an accidental by-product of selection for some other phenotype (BÉGUERET et al. 1994). Explicit population genetic models can help clarify the issue by predicting the patterns of polymorphism expected under different hypotheses about selection. Some features of the het system, such as large numbers of vegetative compatibility groups (VCGs), can be recreated with a model of neutral evolution and a high mutation rate (NAUTA and HOEKSTRA 1996). Trans-species polymorphism, extending >30 million years into the past (WU et al. 1998), however, cannot be explained by a neutral model without postulating extremely large population sizes, which would imply trans-species polymorphism at loci other than het loci as well. Trans-species polymorphism is one of the most striking outcomes of balancing selection in the known recognition systems MHC and SI. In this article, we assume that balancing selection on het loci results from their effect on vegetative incompatibility and find the conditions under which trans-species polymorphism results from this kind of selection.

The exact mode by which balancing selection might be acting at het loci in nature is not clear. The two observed phenotypes are (i) viable heterokaryon formation between genetically similar individuals (identical at all relevant het loci) and (ii) inviability of heterokaryons resulting from genetic difference at a het locus. If there is balancing selection, the failure to form a viable heterokaryon must be favored over the ability to form a viable heterokaryon. Heterokaryon inviability due to vegetative incompatibility has been associated with a reduced risk of transmission of infectious cytoplasmic elements, such as virus-like dsRNAs and senescence plasmids (DEBETS et al. 1994; VAN DIEPENINGEN et al. 1997), which are transmitted via hyphal fusion, and in a reduction in exploitation by aggressive genotypes (DEBETS and GRIFFITHS 1998). Vegetative incompatibility has also been postulated to play a role in limiting outbreeding in certain species (ESSER and BLAICH 1994). Some of

these hypothesized disadvantages of heterokaryon formation have been incorporated into theoretical models of vegetative incompatibility (Hartl *et al.* 1975; Nauta and Hoekstra 1994, 1996; Liu *et al.* 2000). Hartl *et al.* (1975) considered a deterministic two-locus model, with one locus controlling incompatibility and the other (nuclear) locus providing a means for selection. Nauta and Hoekstra (1994) extended this deterministic model to multiple incompatibility loci and cytoplasmically transmitted selective elements and later to a stochastic model with mutation and loss of new VCGs (Nauta and Hoekstra 1996). They derived an expression for the number of VCG types expected at equilibrium in an asexual population and used simulations to study the dynamics of VCG type in populations that had varying degrees of recombination. Transfer of a nongenetic cytoplasmic element between individuals of the same compatibility type was studied by Liu *et al.* (2000) in a spatially explicit model based on the infection of chestnut blight fungus (*C. parasitica*) by hypoviruses.

In this article, we propose a general model of heterokaryon formation in which fusions between individuals identical at all *het* loci result in a loss in fitness to the individuals involved. This model of selection ("compatibility selection"; Nauta and Hoekstra 1996) is an approximation to more detailed models and is roughly equivalent to replacing the disadvantages experienced by particular fusions by the average of such disadvantage in the population. We assume linkage equilibrium among *het* loci. Restricted recombination, in the form of asexual reproduction, has previously been shown to have a minor effect on the number of VCGs maintained in a population, unless the percentage of the population reproducing asexually is very large (Nauta and Hoekstra 1996). An additional effect of asexual reproduction is an increase in the number of encounters that a particular haploid genotype has with others during its "lifetime." Liu *et al.*'s (2000) simulation study found that virus transmission was increased in asexual, highly spatially structured populations of fungi, in which individuals often encountered others of the same compatibility type. We account for the possibility of multiple vegetative encounters between sexual generations in our model.

THE MODEL

Diallelic loci: We illustrate our method in a simple context first and restrict our analysis initially to loci at which only two alleles are possible. We assume that a total of l unlinked loci can affect vegetative incompatibility and that at a given time i of them are polymorphic. Selection is assumed to be sufficiently strong that the two alleles at each polymorphic locus are equally frequent. In our analysis, i is treated as a random variable that increases when a previously monomorphic locus

becomes polymorphic and decreases when one of the polymorphic loci becomes monomorphic. We use the strong selection, weak mutation (SSWM) approximation introduced by Gillespie (1984). The SSWM approximation assumes that mutation is sufficiently infrequent that the time intervals between the appearances of successful mutations are very long and that selection is sufficiently strong that a successful mutation reaches its equilibrium frequency instantaneously on the timescale set by the interval between mutations. The SSWM approximation has proved very useful in the analysis of other models of balancing selection (Sasaki 1989, 1992; Slatkin and Muirhead 1999; Slatkin 2000) because it reduces a highly multiallelic model to a simple Markov chain for the numbers of alleles of different type. In the analysis here, the state variable of the Markov chain is i, the number of dimorphic loci.

Assuming linkage equilibrium, the probability that two randomly chosen haploid individuals have the same genotype is $\frac{1}{2}^i$ and the probability that they differ in genotype is $1 - \frac{1}{2}^i$. If they have the same genotype, the two individuals undergo vegetative fusion that with probability s causes each individual to die or fail to reproduce (for example, because of the transfer of deleterious agents such as mycoviruses). We assume that each individual encounters E other individuals during its lifetime and that each of the E encounters is with an individual chosen randomly and independently from the population. The probability of an individual surviving a single encounter is $1 - \frac{1}{2}^i + (1 - s)/2^i = 1 - s/2^i$ and the probability of it surviving E independent encounters is $w = (1 - s/2^i)^E$, which is $\sim 1 - Es/2^i$ if $Es/2^i$ is small, as it will be in applications of this theory. The quantity w represents the average fitness of the population at equilibrium.

To employ the SSWM approximation, we first consider the fate of a new mutation at one of the $l - i$ monomorphic loci. An individual carrying such a mutation will not form a stable heterokaryon with any other individual and hence will have a relative fitness of 1. While such a mutation is in very low frequency, the deterministic rate of increase in frequency is $\sim 1 - w$. Using the SSWM approximation, then, $2(1 - w) = Es/2^{i-1}$ is the probability in a finite population that the rare mutation will increase in frequency, creating an additional polymorphic locus and increasing i by 1. In a population containing N individuals, the number of mutations at previously monomorphic loci per generation is, on average, $Nu(l - i)$, where u is the per locus mutation rate to alleles affecting compatibility type. Therefore, the net probability per generation that a mutant appears that will result in an increase in i by 1 is

$$\lambda(i) = \frac{Su(l - i)}{2^{i-1}} \tag{1}$$

for $i \geq 0$, where $S = NsE$ is a composite parameter that

plays the role of the scaled selection intensity in models of overdominance.

Stochastic loss of one allele at a diallelic locus results in a decrease in i by 1. To find the rate of stochastic loss, we assume that an allele A at one of the i diallelic loci has a frequency x but that the frequencies at the other $i - 1$ polymorphic loci are $\frac{1}{2}$. We can approximate the expected change in x, $a(x)$, (following the notation of EWENS 1979) under selection alone. The marginal fitness of an individual with A is $\sim w_A = 1 - xsE/2^{i-1}$ and the marginal fitness of an individual with the other allele, a, is $\sim w_a = 1 - (1 - x)sE/2^{i-1}$. Therefore the deterministic change in x under one generation of selection is approximately

$$a(x) = \frac{xw_A}{xw_A + (1 - x)w_a} - x \approx fx(1 - x)\left(\frac{1}{2} - x\right), \quad (2)$$

where $f = sE/2^{i-2}$ is the intensity of balancing selection at each locus. The variance in the change in x in one generation is $b(x) = x(1 - x)/N$ for haploids (EWENS 1979). The expected time to either fixation or loss of A from an initial frequency of $\frac{1}{2}$ is, from EWENS (1979),

$$\bar{t}(\tfrac{1}{2}) = 2\int_0^{1/2} \frac{1}{b(x)\psi(x)}\left(\int_0^x \psi(y)\, dy\right) dx, \quad (3)$$

where

$$\psi(y) = \exp\left[-2\int^y \left(\frac{a(z)}{b(z)}\right) dz\right] = \exp\left(\frac{S(1/2 - y)^2}{2^{i-2}}\right). \quad (4)$$

The integral in Equation 3 cannot be expressed in closed form but it can be approximated very well by

$$\bar{t}(\tfrac{1}{2}) \approx \frac{Ne^{S/2^i}}{4}\sqrt{\frac{2^{3i}\pi}{S^3}} \quad (5)$$

when S is large. The approximation in (5) is similar to that used by TAKAHATA (1990) and SASAKI (1992), but differs slightly. In this case, both fixation and loss of A are possible and must be considered because both result in monomorphism. In Equation 5, the exponential dependence on $S/2^i$ dominates the expression on the right-hand side, so $\bar{t}(\tfrac{1}{2})$ decreases with increasing i.

Continuing with the SSWM approximation, we assume that loss or fixation of an allele occurs in a single generation, and hence $1/\bar{t}(\tfrac{1}{2})$ is the probability per generation that a diallelic locus will become monomorphic. Each of the i diallelic loci can become monomorphic so

$$\mu(i) = \frac{i}{\bar{t}(1/2)} = \frac{2ie^{-S/2^i}}{N}\sqrt{\frac{S^3}{2^{3i}\pi}} \quad (6)$$

is the net rate at which i decreases ($i \geq 1$).

Equations 1 and 6 define the Markov chain. This chain has a particularly simple form and its stationary distribution can be found explicitly. The probability that i takes a particular value is α_i ($0 \leq i \leq l$), which is given by

$$\alpha_i = \alpha_0 \frac{\prod_{j=0}^{i-1}\lambda(j)}{\prod_{j=1}^{i}\mu(j)}, \quad (7)$$

where α_0 is chosen so that $\Sigma_{i=0}^{l} \alpha_i = 1$. There are three parameters: l, the number of loci at which nonself mutations can occur; S, the scaled selection intensity; and $M = Nu$, the scaled mutation rate. For the SSWM approximation to be reasonable, we have to assume that S is large and that M is small.

Numerical analysis of (7) shows that for given values of S and M, there is a critical value of l that distinguishes cases in which the number of loci is limiting from cases in which a balance is achieved between mutation and stochastic loss. The critical value of l, l^*, is the value for which $\lambda(l - 1) = \mu(l)$. Figure 1 shows that l^* is approximately a linear function of $\log_{10}(S)$ and is nearly independent of M. If $l < l^*$, $\lambda(i - 1) > \mu(i)$ for all $i < l$, which means that the rate of gain of diallelic loci always exceeds the rate of stochastic loss. In that case, the stationary distribution is piled up at l implying that all available loci are dimorphic. If a locus becomes monomorphic, it will remain so for only a relatively short time. The biological interpretation is that the number of loci at which mutations can create new VCGs is smaller than the equilibrium number that would otherwise be maintained.

If $l > l^*$, then the stationary distribution of i is centered roughly at the value of i that solves the zero flux condition, $\lambda(i) = \mu(i)$, which can be simplified to

$$M(l - i) = ie^{S/2^i}\sqrt{\frac{S}{2^i\pi}}. \quad (8)$$

In this case, the number of loci is not limiting. When S is large, the stationary distribution is quite sensitive to changes in l only when l is near l^*, as is shown in Figure 2.

More than two alleles per locus: It is straightforward to include loci with more than two alleles. Of the l loci that can affect heterokaryon formation, c_i carry i alleles at any one time, so that the population consists of c_1 monomorphic loci, c_2 dimorphic, and so on, with a maximum of k alleles at any locus. We construct a Markov chain on the set of integers, $c = \{c_1, \ldots, c_k\}$, $\Sigma_{j=1}^{k}c_j = l$. Selection is assumed to be strong enough that allele frequencies at an i-morphic locus are $1/i$. Two randomly chosen individuals have the same genotype with probability $F = \prod_{j=1}^{k}(1/j)^{c_j}$. The probability of surviving a single encounter is $1 - sF$ and the probability of surviving E independent encounters is $\sim 1 - EsF = w$.

As in the model of diallelic loci, each new mutant has a relative fitness of 1 because an individual carrying it cannot form a stable heterokaryon with any other individual in the population. Hence, each new mutant has a selective advantage of $1 - w$ over alleles already present at their equilibrium frequencies. Thus, the probability that a new mutation at a locus with $i - 1$ alleles increases in frequency to become common is $\sim 2EsF$. With c_{i-1} such loci present, the probability that this event occurs is
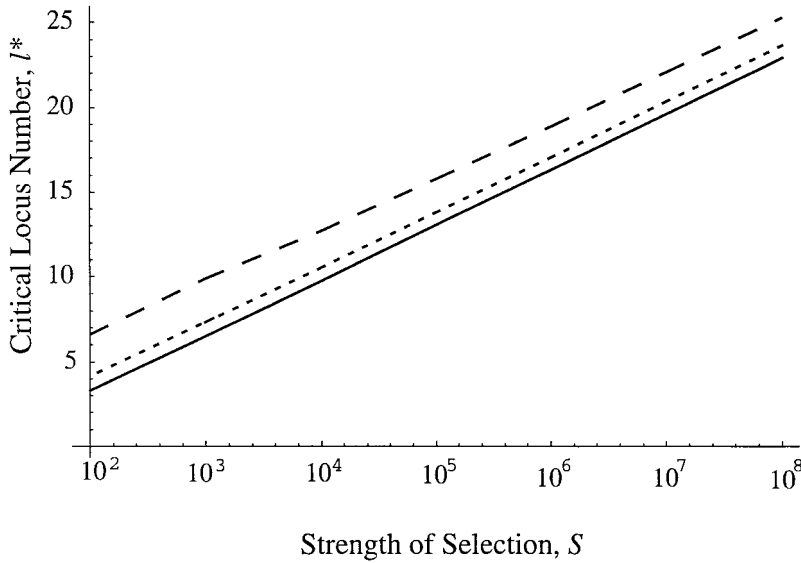
FIGURE 1.—Limiting number of loci ($l*$) as a function of S in the diallelic model ($k = 2$), for three values of M, 1.0 (– – –), 0.01 (- - -), and 0.0001 (—).

$$\lambda_i(\textbf{c}) = c_{i-1}u2SF \qquad (9)$$

(*cf.* Equation 1). This event results in an increase in c_i by one and a decrease in c_{i-1} by one.

The rate of stochastic loss of alleles at a locus with $i > 2$ alleles is slightly different from that for a diallelic locus because only loss and not fixation of an allele must be considered. The mathematical problem is the same as that analyzed previously by SASAKI (1989), TAKAHATA (1990), and SLATKIN and MUIRHEAD (1999). We consider an allele A at frequency x and assume that the other $(i - 1)$ alleles are in frequency $(1 - x)/(i - 1)$. We also assume that allele frequencies at other polymorphic loci are at their deterministic equilibria. The marginal fitness of A is $w_A = 1 - sixEF$ and the marginal fitness of all of the other alleles is $w_a = 1 - si((1 - x)/(i - 1))EF$. Therefore, the average change in x per generation is

$$a(x) = \frac{xw_A}{xw_A + (1 - x)w_a} - x \approx x(1 - x)\left(\frac{1}{i} - x\right)\frac{i^2 sEF}{(i - 1)}, \qquad (10)$$

the variance is $b(x) = x(1 - x)/N$, and

$$\psi(y) = \exp\left(SF(1/i - y)^2\left(\frac{i^2}{i - 1}\right)\right) \qquad (11)$$

(*cf.* Equation 4).

If the coefficient of $(1/i - y)^2$ in (11) is large, then the ultimate loss of A is essentially certain; the probability of fixation of A is proportional to e^{-SF}. In that case, the time to loss is most easily found by ignoring the possibility of fixation and using Equations 4.39 and 4.40 from EWENS (1979) for the case in which only loss of A is possible. The result is, approximately,

$$\bar{t}(1/i) = Ne^{SF/(i-1)}\sqrt{\frac{\pi(i - 1)^3}{S^3F^3i^2}} \qquad (12)$$

and hence the probability per generation of a loss of an allele at a locus with i alleles is

$$\mu_i(x) = \frac{c_i i e^{-SF/(i-1)}}{N}\sqrt{\frac{S^3F^3}{\pi(i - 1)^3}}, \qquad (13)$$

which is the rate at which c_i is reduced by 1 and c_{i-1} is increased by 1.

Equations 9 and 13 provide the transition probabilities for the Markov chain. There is no absorbing state so the chain has a stationary distribution, but the chain is not a continuant, so there is no analytic expression for the stationary distribution, as there is for the model of diallelic loci (Equation 7). The zero flux conditions, $\lambda_i(\textbf{c}) = \mu_i(\textbf{c})$, have to be solved numerically.

As in the case of the model of diallelic loci, the results depend on l and the two composite parameters M and S. The maximum number of VCGs possible in a population,
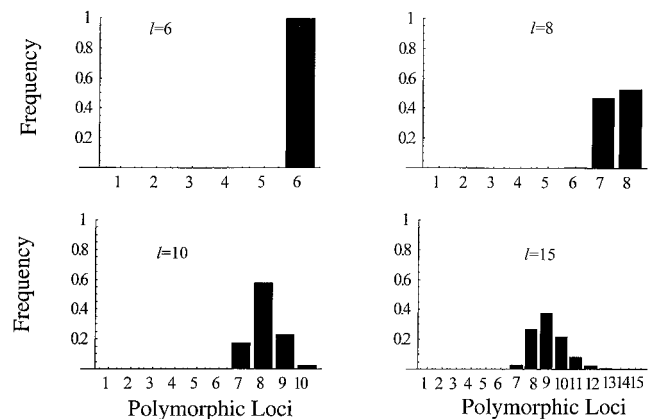

FIGURE 2.—Stationary distributions of i, the number of polymorphic loci, for the diallelic model. For these parameters ($M = 0.1$, $S = 10^3$), $l* = 8$.
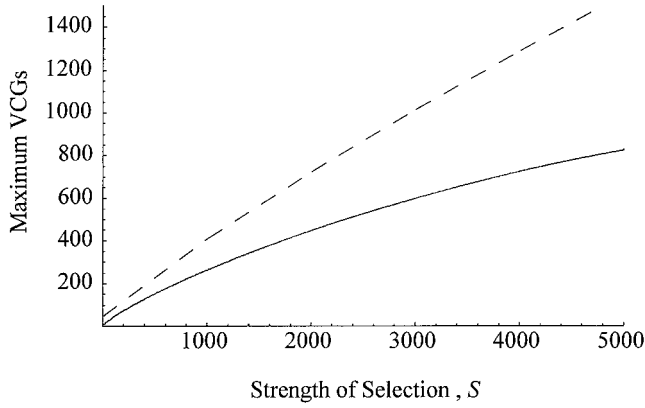
FIGURE 3.—Maximum number of vegetative compatibility groups (VCGs) at equilibrium for diallelic (—) and triallelic loci (– – –). $M = 0.1$, S varies from 10 to 5000.

FIGURE 4.—Equilibrium numbers of mono-, di-, to k-morphic loci, for different combinations of M and S ($M = 0.1$ and 0.01; $S = 10^3$ and 10^5).

assuming equal frequencies of alleles at each locus and linkage equilibrium between loci, is $\prod_{i=2}^{k} i^{c_i}$, with an absolute maximum of k^l if all loci were polymorphic to the maximum amount possible. As shown in Figure 3, the model predicts large numbers of VCGs for reasonable values of M and S, even for small values of k.

As in the previous section, there is a critical number of loci, l^*, that determines the behavior of the model. The value of l^* is the largest value of l for which $\lambda_i(0, \ldots 0, 1, l-1) > \mu_i(0, \ldots 0, l)$. If $l < l^*$, the rate of appearance of new mutations exceeds the rate of loss even when every locus has the maximum possible number of alleles, which means that the number of loci and hence the total number of VCGs are limited by the number of loci at which such mutations can occur. For given values of M and S, l^* decreases as k increases. For example, with $M = 0.01$ and $S = 10^5$, $l^* = 10$ for the diallelic ($k = 2$) model, $l^* = 6$ for $k = 3$, and $l^* = 2$ for $k = 10$. When $l < l^*$, most loci have the maximum number of alleles. But for $k > 3$, we find this can occur only when the rate of mutation to new specificities is very high ($M > 1$).

If $M < 1$, even extremely strong selection does not necessarily result in highly polymorphic loci if $l > l^*$ (Figure 4). For most of Figure 4, there are essentially no loci with more than three alleles. Some multiallelic loci can be maintained with the higher mutation rate shown, but even with $M = 0.1$, $S = 10^5$, most loci (6.1 out of 10) have three or fewer alleles. With greater numbers of available loci, the conditions for multiallelism become even stricter. Increasing the mutation rate tends to increase overall diversity by increasing the number of polymorphic loci, rather than by increasing the number of alleles per locus.

We conclude that the behavior of the model with multiallelic loci is similar to that of the model with only diallelic loci. Two general features emerge.

i. Loci will be equally polymorphic to the extent possible. Only when all loci are diallelic are a significant number of triallelic loci maintained, and more than three alleles
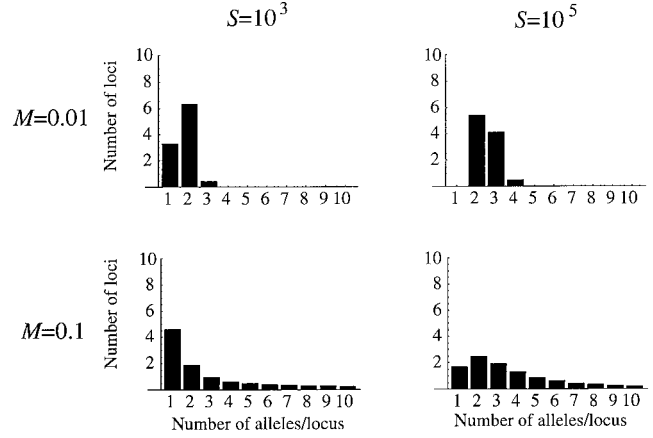
per locus require very strong selection or very small numbers of available loci.

ii. The behavior of the model depends on whether the number of loci (l) is limiting. If l is small, then all available loci will be polymorphic to the maximum possible extent.

Trans-species polymorphism: For trans-species polymorphism to be detected, alleles must remain in the population for very long times, much longer than alleles at neutral or weakly selected loci. In a haploid species, the average time until loss of alleles initially present in a moderate frequency is $\sim N$ generations. Trans-species polymorphism is possible if $\bar{t}/N \gg 1$, where $\bar{t}$ is the expected time to loss of an allele initially present at its equilibrium frequency ($\frac{1}{2}, \frac{1}{3}, \ldots$). For the diallelic model, an analytical expression for the average age of an allele is given by (3).

Average allele age for the multiple-allele model may be calculated by construction of another Markov chain, using the rates of allele gain and loss derived above. In this Markov chain, the state is the current number of alleles at a locus containing an allele of interest. The state changes as alleles are gained and lost at the locus until eventually either the locus becomes monomorphic or the allele of interest itself is lost from the population. The average age of an allele, given that it starts in state i, $i > 1$, is

$$\bar{t}_i = \sum_{j=1}^{k} P_{ij}\bar{t}_j + 1, \tag{14}$$

where P_{ij} is the probability that a locus with i alleles will have j alleles at the next time step (Ewens 1979). These probabilities are easily calculated using (9) and (13). We are concerned only with the time that an allele is at a polymorphic locus, so that $\bar{t}_1 = 0$ and $P_{11} = 1$. The values of $\bar{t}_2 \ldots \bar{t}_k$ have to be found numerically, and the mean age of alleles found at a polymorphic locus is calculated as

$$\bar{t} = \frac{\sum_{i=2}^{k} \bar{t}_i c_i}{(1 - c_1)}. \qquad (15)$$

The quantity $\bar{t}/N$ depends strongly on the scaled mutation parameter M and on whether the number of loci is limiting. If $l > l^*$, alleles persist for only slightly longer times than neutral alleles but if $l < l^*$ persistence times can be dramatically longer. For example, in the diallelic case, with $M = 0.1$ and $S = 10^3$ ($l^* = 8$), $\bar{t}/N = 43{,}825.9$ for $l = 6$, 7.6 for $l = 8$, 2.7 for $l = 10$, and 1.4 for $l = 15$, which are the values for which the stationary distributions are shown in Figure 2. This pattern is typical and leads us to the conclusion that for trans-species polymorphism to be possible when $k = 2$, l must be $<l^*$, implying that all or nearly all of the available loci are polymorphic at any time. There is still turnover of alleles but only on a very long timescale. This conclusion is also true if $k > 2$, but in this case, biologically plausible values of l (5–15 for most species studied) are generally $>l^*$ (Table 1). If $k > 2$ and l is on the order of 10, alleles turn over more slowly than neutral alleles but not slowly enough to account for trans-species polymorphism. In most cases, average allele age is only slightly greater than in the neutral case (Table 1) and is far less than is usually found with balancing selection.

If $k > 2$, trans-species polymorphism is possible only if M is very small. The equilibrium level of polymorphism is largely determined by S, l, and k, but the turnover of alleles is driven by mutation. Lower values of M result in longer persistence times. We conclude, then, that trans-species polymorphism appears to require severe limitation of mutational opportunity, through a very low mutation rate, or through some constraints on the number of possible alleles at a locus, or both.

DISCUSSION

We have investigated a model that contains the essential features of vegetative incompatibility of the type found in *N. crassa* and many other species of filamentous fungi. The model's parameters are (i) the number of loci at which mutations creating new incompatibility types can occur (l), (ii) the number of alleles possible at each of these loci (k), (iii) the mutation rate scaled by the effective population size ($M = Nu$), and (iv) the scaled selection intensity ($S = NsE$, where E is the number of opportunities an individual has for vegetative fusion).

There are some constraints on the values of these parameters. For the model to represent strong stabilizing selection, S must be $\geqslant 1$. It is also biologically reasonable that M be $\leqslant S$. Genetic studies cited in the Introduction indicate that the biologically interesting range for l is 5–15.

Our results show that, as long as selection is strong (*i.e.*, $S \geqslant 1$), genetic polymorphism will be maintained at loci affecting vegetative incompatibility and numerous incompatibility types will be found under a wide range of values of the other parameters. The model predicts that as many

TABLE 1

Scaled average allele age, $\bar{t}/N$, for different mutation rates (M), selection strength (S), locus polymorphism (k), and number of loci (l)

		Maximum alleles per locus, k		
	No. of loci, l	2	3	10
$M = 0.01$, $S = 10^4$	10	371.7	11.9	10.5
	20	6.0	4.3	4.1
$M = 0.0001$, $S = 10^4$	10	4384.1	841.2	835.2
	20	255.6	214.4	214.3

Ages are much greater than the neutral expectation when mutation is limited, either by a low mutation rate or by limitations on the number of loci and possible alleles.

loci as possible will be polymorphic and that, if the number of loci is not limited, most of the polymorphic loci will have only two alleles. If the number of loci is limited, then all available loci will be polymorphic and some may carry more than two functionally different alleles per locus. To the extent possible, loci will tend to have equal numbers of alleles.

The symmetry of our results and in particular the prediction of equal numbers of alleles per locus is a consequence of the symmetry of the underlying assumptions. Mutation rates and numbers of possible alleles could easily differ among loci, leading to differences in the numbers of alleles observed.

Trans-species polymorphism is not a necessary consequence of selection of this type, as it is for strong balancing selection affecting a single locus. The reason is that, although polymorphism can be maintained at loci responsible for vegetative incompatibility, considerable turnover of alleles can occur. As a result, no allele persists long enough for trans-species polymorphism to be commonly found.

Our model predicts that trans-species polymorphism will be found if the rate of turnover of alleles is constrained in some way. One possibility is that the rate of mutation to functionally different alleles at each locus is very small. That would result in long persistence times because only when a new allele appears is there pressure for existing alleles to be lost by genetic drift. Another possibility is that there are so few loci and so few alleles per locus that, at equilibrium, nearly all possible alleles are present. In that case, few if any new functionally different alleles could enter the population and cause turnover of existing alleles.

On the basis of available information, low mutation rate is a more plausible explanation for observations of trans-species polymorphism than is exhaustion of available alleles. At *het-c* in *N. crassa* and related species, alleles exhibiting trans-species polymorphism differ by an insertion/deletion (indel) motif, a mutational event that is much rarer than nucleotide substitutions or the insertion

or deletion of only one or two nucleotides. That observation is consistent with the idea that the mutation rate to functionally different alleles is very low. Under laboratory conditions, it was possible to generate a limited number of new *het-c* allelic specificities by altering the indel pattern in the specificity domain of *het-c* (Wu and Glass 2001). Because synthetic functionally different alleles have been created, it does not appear as if all possible alleles are present, at least at *het-c*.

If trans-species polymorphisms of the type found by Saupe and Glass (1997) and by Wu *et al.* (1998) are not the result of their role in vegetative incompatibility, an alternative explanation is that those alleles directly experience balancing selection because of some pleiotropic effect. Additional evidence of trans-species polymorphism from other loci involved in vegetative incompatibility will help resolve this issue.

LITERATURE CITED

Anagnostakis, S. L., B. Hau and J. Kranz, 1986 Diversity of vegetative compatibility groups of *Cryphonectria parasitica* in Connecticut [USA] and Europe. Plant Dis. **70:** 536–538.

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

Arden, B., and J. Klein, 1982 Biochemical comparison of major-histocompatibility-complex-molecules from different subspecies of *Mus musculus*—evidence for trans-specific evolution of alleles. Proc. Natl. Acad. Sci. USA **79:** 2342–2346.

Bégueret, J., B. Turcq and C. Clave, 1994 Vegetative incompatibility in filamentous fungi: *het* genes begin to talk. Trends Genet. **10:** 441–446.

Bjorkman, P. J., and P. Parham, 1990 Structure, function, and diversity of Class I major histocompatibility complex molecules. Ann. Rev. Biochem. **59:** 253–288.

Clarke, A. E., and E. Newbigin, 1993 Molecular aspects of self-incompatibility in flowering plants. Annu. Rev. Genet. **27:** 257–279.

Cortesi, P., and M. G. Milgroom, 1998 Genetics of vegetative incompatibility in *Cryphonectria parasitica*. Appl. Environ. Microbiol. **64:** 2988–2994.

Debets, A. J. M., and A. J. F. Griffiths, 1998 Polymorphism of *het* genes prevents resource plundering in *Neurospora crassa*. Mycol. Res. **102:** 1343–1349.

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

Esser, K., and R. Blaich, 1994 Heterogenic incompatibility in fungi, pp. 211–232 in *The Mycota, I: Growth, Differentiation and Sexuality*, edited by J. G. H. Wessels and F. Meinhardt. Springer-Verlag, New York.

Ewens, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, New York.

Figueroa, F., E. Guenther and J. Klein, 1988 MHC polymorphism predating speciation. Nature **335:** 265–267.

Gillespie, J. H., 1984 Some properties of finite populations experiencing strong selection and weak mutation. Am. Nat. **121:** 691–708.

Glass, N. L., and G. A. Kuldau, 1992 Mating type and vegetative incompatibility in filamentous Ascomycetes. Annu. Rev. Phytopathol. **30:** 201–224.

Glass, N. L., D. J. Jacobson and P. K. T. Shiu, 2000 The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. Annu. Rev. Genet **34:** 165–186.

Hartl, D. L., E. R. Dempster and S. W. Brown, 1975 Adaptive significance of vegetative incompatibility in *Neurospora crassa*. Genetics **81:** 553–569.

Howlett, B., J. F. Leslie and D. D. Perkins, 1993 Putative multiple alleles at the vegetative (heterokaryon) incompatibility loci *het-c* and *het-8* in *Neurospora crassa*. Fungal Genet. Newsl. **40:** 40–42.

Ioerger, T. R., A. G. Clark and T. H. Kao, 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA **87:** 9732–9735.

Jinks, J. L., C. E. Caten, G. Simchen and J. H. Croft, 1966 Heterokaryon incompatibility and variation in wild populations of *Aspergillus nidulans*. Heredity **21:** 227–239.

Kelly, J. K., and M. J. Wade, 2000 Molecular evolution near a two-locus balanced polymorphism. J. Theor. Biol. **204:** 83–101.

Klein, J., 1980 Generation of diversity at MHC loci: implications for T cell receptor repetoires, pp. 239–253 in *Immunology 80*, edited by M. Fogereau and J. Dausset. Academic Press, London.

Klein, J., A. Sato, S. Nagl and C. O'hUigin, 1998 Molecular trans-species polymorphism. Annu. Rev. Ecol. Syst. **29:** 1–21.

Leslie, J. F., 1993 Fungal vegetative compatibility. Annu. Rev. Phytopath. **31:** 127–150.

Liu, Y.-C., R. Durrett and M. G. Milgroom, 2000 A spatially-structured stochastic model to simulate heterogenous transmission of viruses in fungal populations. Ecol. Model. **127:** 291–308.

Milgroom, M. G., and P. Cortesi, 1999 Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. Proc. Natl. Acad. Sci. USA **96:** 10518–10523.

Mir-Rashed, N., D. J. Jacobson, M. R. Dehghany, O. C. Micali and M. L. Smith, 2000 Molecular and functional analyses of incompatibility genes at *het-6* in a population of *Neurospora crassa*. Fungal Genet. Biol. **30:** 197–205.

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

Nauta, M. J., and R. F. Hoekstra, 1994 Evolution of vegetative incompatibility in filamentous ascomycetes. I. Deterministic models. Evolution **48:** 979–995.

Nauta, M. J., and R. F. Hoekstra, 1996 Vegetative incompatibility in ascomycetes: highly polymorphic but selectively neutral? J. Theor. Biol. **183:** 67–76.

Pandit, A., and R. Maheshwari, 1996 A demonstration of the role of *het* genes in heterokaryon formation in *Neurospora* under simulated field conditions. Fungal Genet. Biol. **20:** 99–102.

Perkins, D. D., and R. H. Davis, 2000 *Neurospora* at the millennium. Fungal Genet. Biol. **31:** 153–167.

Pontecorvo, G., 1956 The parasexual cycle in fungi. Annu. Rev. Microbiol. **10:** 393–400.

Powell, A. J., D. J. Jacobson and D. O. Natvig, 2001 Allelic diversity at the *het-c* locus in *Neurospora tetrasperma* confirms outcrossing in nature and reveals an evolutionary dilemma for pseudohomothallic ascomycetes. J. Mol. Evol. **52:** 94–102.

Sasaki, A., 1989 Evolution of pathogen strategies. Ph.D. Thesis, Kyushu University, Fukuoka, Japan.

Sasaki, A., 1992 The evolution of host and pathogen genes under epidemiological interaction, pp. 247–263 in *Population Paleo-Genetics*, edited by N. Takahata. Japan Scientific Society Press, Tokyo.

Saupe, S. J., 2000 Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. Microbiol. Mol. Biol. Rev. **64:** 489–502.

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

Saupe, S. J., C. Clave and J. Begueret, 2000 Vegetative incompatibility in filamentous fungi: *Podospora* and *Neurospora* provide some clues. Curr. Opin. Microbiol. **3:** 608–612.

Skupski, M. P., D. A. Jackson and D. O. Natvig, 1997 Phylogenetic analysis of heterothallic *Neurospora* species. Fungal Genet. Biol. **21:** 153–162.

Slatkin, M., 2000 Balancing selection at closely linked, overdominant loci in a finite population. Genetics **154:** 1367–1378.

Slatkin, M., and C. A. Muirhead, 1999 Overdominant alleles in a population of variable size. Genetics **152:** 775–781.

Takahata, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87:** 2419–2423.

van Diepeningen, A. D., A. J. M. Debets and R. F. Hoekstra, 1997

Heterokaryon incompatibility blocks virus transfer among natural isolates of black *Aspergilli*. Curr. Genet. **32:** 209–217.

VEKEMANS, X., and M. SLATKIN, 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. Genetics **137:** 1157–1165.

WORRALL, J. J., 1997 Somatic incompatibility in basidiomycetes. Mycologia **89:** 24–36.

WRIGHT, S., 1939 The distribution of self-sterility alleles in populations. Genetics **24:** 538–552.

WU, J., and N. L. GLASS, 2001 Identification of specificity determinants and generation of alleles with novel specificity at the *het-c* heterokaryon incompatibility locus of *Neurospora crassa*. Mol. Cell. Biol. **21:** 1045–1057.

WU, J., S. J. SAUPE and N. L. GLASS, 1998 Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. Proc. Natl. Acad. Sci. USA **95:** 12398–12403.