Importance of the Life Cycle in Sympatric Host Race Formation and Speciation of Pathogens

Tatiana Giraud, Lorys M. M. A. Villaréal, Frédéric Austerlitz, Mickaël Le Gac, and Claire Lavigne

Ecologie, Systématique et Evolution, UMR 8079 CNRS-Université Paris Sud, Bâtiment 360, Université Paris-Sud, F-91405 Orsay cedex, France.
Accepted for publication 5 October 2005.

ABSTRACT


Numerous morphological species of pathogenic fungi have been shown to actually encompass several genetically isolated lineages, often specialized on different hosts and, thus, constituting host races or sibling species. In this article, we explore theoretically the importance of some aspects of the life cycle on the conditions of sympatric divergence of host races, particularly in fungal plant pathogens. Because the life cycles classically modeled by theoreticians of sympatric speciation correspond to those of free-living animals, sympatric divergence of host races requires the evolution of active assortative mating or of active host preference if mating takes place on the hosts. With some particular life cycles with restricted dispersal between selection on the host and mating, we show that divergence can occur in sympatry and lead to host race formation, or even speciation, by a mere process of specialization, with strong divergent adaptive selection. Neither active assortative mating nor active habitat choice is required in these cases, and this may explain why the phylo-genetic species concept seems more appropriate than the biological species concept in these organisms.

Additional keywords: host choice, natural selection against immigrants from divergent habitats, parasites.

Many morphological species of fungal pathogens have been shown to actually encompass several genetically isolated lineages, recognized through their intersterility (2,10,21) or through molecular evidence of restricted gene flow (4,14,33,41). These lineages often are specialized on different plant species and, therefore, are called host races or formae specialis when some restricted but significant gene flow remains, and sibling species when the lineages are evolutionary independent (13) (i.e., when gene flow occurs at a low enough level to allow divergence by selection and genetic drift). Host-specific divergence of parasites may evolve as a consequence of limited dispersal or of trade-offs in adaptation (50). Some parasites may have a restricted host range simply because they do not come in contact with other host species, corresponding to allopatric differentiation. Alternatively, host specificity may arise because of adaptive specialization (52): existence of trade-offs between adaptation to different hosts or a high cost of being a generalist can lead to host-specific differentiation of parasites in sympatry. Here we will focus on sympatric divergence, which is the most debated process of genetic differentiation. Because some words of the speciation literature are used with different meanings in the fields of evolution and phytopathology, we define the ambiguous terms as used in this article in Table 1.

The idea that strong selection for specialization can cause sympatric divergence is common in the phytopathological literature, but most authors are unaware that this is not sufficient cause for speciation from a theoretical point of view (23). Indeed, the entire literature on sympatric speciation shows how the existence of alternative alleles beneficial in different habitats does not readily lead to genetic isolation in sympatry (22). Strong selection allows maintenance of adaptive polymorphism when a single adaptive locus is involved (35). However, if several loci are involved, optimal adaptation on a given host requires the accumulation of all favorable alleles at all loci within a genotype. This will generate linkage disequilibrium between these different loci. The favorable genetic combinations and linkage disequilibrium will be broken if mating events occur between individuals adapted to different habitats. Moreover, these mating events will allow gene flow at loci not linked to the genes involved in host adaptation (20,46), thus preventing neutral genes or genes involved in postzygotic isolation to become fixed. Therefore, for a polymorphism for host exploitability to lead to true divergence for sexual populations occurring in sympathy, reproductive events between individuals on different hosts must be prevented by assortative mating, which is, however, often difficult to achieve (22). The simplest way to prevent these reproductive events is to have the same gene or genes controlling pleiotropically both fitness and assortative mating (30–32,37,46). However, very few biological cases of such kind of genes acting pleiotropically on fitness and assortative mating have been reported (19). Another way to reduce recombination between two populations specialized on different hosts is to build up an association between host-based fitness genes and either assortative mating genes or host preference genes if mating is restricted on hosts. Several theoretical models examined the conditions that allow evolution of linkage disequilibrium between these different types of genes (16,17,27,44,51).

However, all these theoretical models on sympatric speciation were based on life cycles of free-living animals for which individuals mate at random in the population as a whole or actively choose a host where they mate. This also corresponds to the life cycle of many fungal species (1), except that they cannot choose their host, spores usually being passively dispersed by wind or water. Therefore, sympatric divergence in these fungi is expected to be possible only through mate choice (i.e., active prezygotic isolation or active assortative mating). However, some other fungal pathogens have life-history traits that could generate specific modes of sympatric speciation by host specialization. In particular, some fungal plant pathogens have restricted dispersal between...
development on the host and mating (1), so that individuals mate mostly with other individuals that also were sufficiently adapted to the same host to be able to develop on it. Therefore, with such life cycles, mating between individuals specialized on the same host may be a direct consequence of specialization, requiring neither the evolution of host choice nor active assortative mating. The conditions allowing sympatric divergence by host specialization may be dramatically different depending on the life cycle considered, but this has not yet been explored theoretically.

To this end, we develop here a model to study the conditions that allow sympatric divergence without any active assortative mating gene in pathogens with a life cycle with restricted dispersal between development on the host and mating and with no possibility of active host choice. Our model was patterned after some pathogenic filamentous Ascomycete fungi, such as *Venturia inaequalis*, responsible for scab, the major apple disease in most areas of the world; *V. inaequalis* also attacks crab apple, pyracantha, Hawthorn, various ornamentals of the genus Malus, loquat, and other plants. Asexual conidia and ascospores are wind disseminated and, thus, cannot choose the plant on which they land. Late in the season, the haploid mycelium penetrates deep into leaf tissues and proceeds to form ascocarps (1). Therefore, mating can occur only between individuals that developed inside the same leaf. Zygotes undergo immediate meiosis, giving rise to dispersing haploid ascospores (1). Other examples of specialized parasitic ascomycetes that reproduce on their host and have gametes with low dispersal ability include the agent of powdery mildew, *Erysiphe graminis* (1). Parasitic life cycles with mating on the host and restricted gamete dispersal also are found in some other taxa, for instance in Oomycota. Specialized species exist, for instance, in the genera Phytophthora or Peronospora, which are responsible for important crop diseases such as the late blight of potato. Mycelia grow inter- or intracellularly, where they produce gametes, so that sexual reproduction should occur only between strains infecting the same host individual (1).

Thus, we considered a population of a parasite species living in an environment with two different hosts and having the following life cycle: (i) individuals undergo development in one host and stay in this host for mating, (ii) mating takes place at random within each host, and (iii) offspring disperse at random between the two hosts. A set of loci controls the host-based fitness but there are no loci controlling active habitat preference or active assortative mating. Divergence between populations with such a life cycle is truly sympatric because offspring disperse at random between the two hosts and, consequently, the probability of mating between two individuals depends only on their genotypes (at fitness genes), because only individuals that could develop on a particular host can reproduce on it (29).

We developed two versions of the model: a deterministic model for the simplest cases (one and two loci), and an individual-based stochastic model for the other cases. We examined the intensities of selection that allow evolution of linkage disequilibria between the different fitness genes and a significant reduction in gene flow between the two fungal populations adapting to the two hosts. We also examine whether these conditions similarly allow sympatric divergence with the alternative life cycle (i.e., with unrestricted gamete dispersal between selection on the host and mating, or with significant but restricted gamete dispersal).

### THEORY AND APPROACHES

**General description.** We considered a haploid pathogenic species with discrete generations, living in an environment consisting of two hosts, $\alpha$ and $\beta$. The two hosts were equally abundant and the same quantity of parasite could develop on each of them. Thus, we assumed soft selection (i.e., fitness affected only the proportions of the different genotypes) (11). We assumed either that all individuals underwent their development in one host and stayed in this host for mating or that a proportion, $m$, of gametes from each host dispersed to the other host to check the effect of dispersal between selection and mating. Mating took place at random within each host. Offspring then dispersed at random between the two hosts.

Our model assumed no active host choice or active assortative mating. A set of loci controlled the fitness in the two hosts (i.e., the probability to infect and develop on the hosts). We constructed two types of models with different hypotheses concerning host based fitness loci (Table 2). In model 1, we considered a number, $L$, of host-based fitness loci, named $\text{A loci}$, each with an allele $A_i$ with a relative fitness $1 + s \text{ host } \alpha$ and $1$ in host $\beta$, and an allele $A'_i$ with, symmetrically, a relative fitness of $1$ in host $\alpha$ and $1 + s$ in host $\beta$ (Table 2). In model 2, we considered two $L$ host-based fitness loci of two types (Table 2): (i) a number, $L$, of $\text{A loci}$, each with an allele $A_i$ with a relative fitness $1 + s$ in host $\alpha$ and $1 – c$ in $\beta$ and (ii) a number, $L$, of $\text{B loci}$, each with an allele $B_i$ with a relative fitness $1 – c$ in host $\alpha$ and $1 + s$ in $\beta$ (Table 2).

### TABLE 1. Definitions used in this article of words with ambiguous meanings

<table>
<thead>
<tr>
<th>Terms used in this paper</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assortative mating</td>
<td>Mating between two individuals carrying the same alleles of specialization</td>
</tr>
<tr>
<td>Active assortative mating</td>
<td>Active choice of a mating partner (i.e., refusal to mate until finding an individual with the same alleles of specialization)</td>
</tr>
<tr>
<td>Host race or formae specialis</td>
<td>Group of individuals specialized on a host species, with restricted but significant gene flow connecting them to groups of individuals specialized on other host species</td>
</tr>
<tr>
<td>Sibling species</td>
<td>Group of individuals specialized on one host species, gene flow with groups of individuals specialized on other host species occurring at a level low enough to allow divergence by selection and genetic drift</td>
</tr>
<tr>
<td>Host exploitability or host exploitation ability</td>
<td>Ability to develop on a given host due to carrying the appropriate alleles of specialization</td>
</tr>
<tr>
<td>Host preference or host choice</td>
<td>Active choice of a host upon which to develop (i.e., active dispersal until finding the host for which one has the appropriate alleles of specialization)</td>
</tr>
<tr>
<td>Host-based fitness genes or fitness genes</td>
<td>Genes with alternative alleles allowing development in different host species (i.e., providing specialization to different host species)</td>
</tr>
<tr>
<td>Preference genes or host choice genes</td>
<td>Genes making an individual actively choose a host upon which to develop (i.e., actively dispersing until finding the appropriate host specified by the allele)</td>
</tr>
<tr>
<td>Assortative mating genes</td>
<td>Genes making an individual actively choose a mating partner (i.e., refuse to mate until finding a partner with the alleles at other loci specified by the allele carried at the assortative mating genes)</td>
</tr>
<tr>
<td>Specialization</td>
<td>Existence of genes with alternative alleles allowing development on different host species, not necessarily implying restricted gene flow between individuals carrying different allele types</td>
</tr>
</tbody>
</table>

### TABLE 2. Fitness of the different alleles on the two hosts ($\alpha$ and $\beta$) for model 1 (one type of locus: the two alleles of each locus are beneficial on alternative hosts) and model 2 (two types of loci: each locus has an allele with the same fitness on both hosts)*

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_i$</td>
<td>$A'_i$</td>
</tr>
<tr>
<td>$A_i$</td>
<td>$A'_i$</td>
</tr>
<tr>
<td>$B_i$</td>
<td>$B'_i$</td>
</tr>
</tbody>
</table>

* $X_i$ and $X'_i$ represent the two alleles of the locus $X_i$, $i$ ranging from 1 to $L$. 

Vol. 96, No. 3, 2006 281
host $\beta$, and an allele $A_i$, with a fitness of 1 in both hosts; and (ii) a number, $L$, of B loci, which have, symmetrically, each an allele $B_i$ with a relative fitness of $1 + s$ in host $\beta$ and a fitness $1 - c$ in host $\alpha$, and an allele $B'$, with a fitness of 1 in both hosts. The parameters $s$ and $c$ are selective coefficients, being real numbers; $s$ can take any positive value and $c$ is between 0 and 1, the values of $s$ and $c$ being independent. All loci within a genome had the same value of $s$, and of $c$ when applicable. We expected divergence between host races to be more difficult to achieve using model 2, in which fitness loci have an allele with the same fitness in both hosts, than using Model 1, in which fitness loci have alleles beneficial on alternative hosts.

There initially was no linkage disequilibrium and all loci were considered to be physically unlinked. We investigated two types of fitness functions in each model, one related to a model of multiplicative interactions among loci and one to an additive model. In model 1 (with only one type of locus), let $X$ be the number of $A_i$ alleles over all loci for a given individual. The fitness of this individual in host $\alpha$ is then

$$(1 + s)^X$$

for the model related to multiplicative interactions

$$(1 + sX^c)$$

for the model related to additive interactions

The parameter $\varepsilon$ represents the degree of epistasis among fitness loci: $\varepsilon = 1$ means no epistasis between loci, and equation 1 then simulates multiplicative interactions among loci and equation 2 additive interactions; $\varepsilon > 1$ means synergistic epistasis between loci, making them act more than multiplicatively or additively; and $\varepsilon < 1$ means antagonistic epistasis between loci, making them act less strongly than multiplicatively or additively.

In model 2 (with two types of loci), let $X$ be the number of $A_i$ alleles over loci (which have a fitness of $1 + s$ on host $\alpha$) and $Y$ the number of $B_i$ alleles over loci (which have a fitness of $1 - c$ on host $\alpha$) of a given individual. The fitness of this individual on host $\alpha$ is then

$$(1 + s)^X(1 - c)^Y$$

for the model related to multiplicative interactions

$$(1 + sX^c - cY^c)$$

for the model related to additive interactions

From these formulae, epistasis in model 2 can occur only among loci of the same type: among $A$ loci on the one hand and among $B$ loci on the other hand. Therefore, its effect could be investigated only for cases with more than two loci of each type ($2L \geq 4$), so only with the stochastic models.

With such fitness functions, increasing the number of loci increases the fitness of the most adapted genotypes, precluding direct comparisons. However, whether increasing the number of loci will facilitate divergence for a given coefficient of selection per locus will depend on the level of linkage disequilibrium that can be built up among these loci. We used these fitness functions to investigate whether a life cycle without dispersal between selection and mating can build up sufficient linkage disequilibrium to allow several fitness genes to combine their effects and, thereby, to promote divergence between host races of parasites.

We developed two versions of the models: (i) a deterministic model for the simplest cases (one and two loci), in which equilibrium values of interhost linkage disequilibrium and gene flow were found by solving analytically the equations describing the changes in genotype frequencies; and (ii) an individual-based stochastic model for more than two loci, in which we examined, by simulations, the intensities of selection that allow evolution of interhost linkage disequilibria and reduction in neutral gene flow. We estimated the possibility of neutral gene exchange by calculating contributions of individuals from one host to the pool of gametes produced in the other host. The codes of the different programs are available on request.

Deterministic model based on genotype frequencies. This version of the model was built for the cases of one locus (model 1) and two loci (models 1 and 2) using the software Mathematica (version 4.1; Wolfram Research, Inc., Champaign, IL) We first computed the exact equations giving the relations between the genotypic frequencies $f_{\beta\alpha\alpha}$ at time $t$ and $t + 1$, where $G$ denotes any of the possible genotypes and $h$ the host (i.e., $h = \alpha$ or $\beta$). Then, we obtained the system of equations for the equilibrium frequencies $f_{\beta\alpha\alpha}$ by saying that at equilibrium $f_{\beta\alpha\alpha} = f_{\alpha\beta\beta} = f_{\beta\alpha\alpha}$ for all $G$ and $h$. This system was solved analytically. Equations and analytical solutions for the $f_{\beta\alpha\alpha}$’s are very lengthy, so we do not show them here; however, they are given in the Mathematica notebooks and available upon request.

Using the $f_{\beta\alpha\alpha}$’s, we computed the probability at equilibrium of gene flow between populations from the two hosts. It was computed as the contribution at equilibrium of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ (i.e., the mean fitness on host $\alpha$ of individuals coming from host $\beta$ relative to the mean fitness on host $\alpha$ of the total population). Indeed, gene flow between the two host races depended on the relative proportions of gametes present on a given host that were produced by individuals coming from the two different hosts. These proportions depended on the relative abilities of the individuals coming from the two hosts to develop and produce gametes on a given host and, thus, was a function of the relative fitness of the two pools of individuals on the host considered. For the two-locus case, we also computed interhost linkage disequilibrium between fitness loci at equilibrium (D’) (36).

Stochastic model based on haplotypes. The stochastic version of the model, individual-based using a matrix of haplotypes, was written with the software Scilab 2.7 (1989–2003; INRIA, available online). In this version of the model, a number, $N$, of haploid individuals (spores) arrived on each host at each generation. $N$ gametes were produced on each host. Each gamete was produced by a given individual present on a host with a probability corresponding to its relative fitness on this host. Crosses took place at random between gametes present on a given host and they gave $N$ haploid offspring in each host. Offspring from each host then dispersed randomly onto the two hosts. The simulations presented in this article were performed by setting $N$ to 200 and the initial allelic frequencies to 0.5.

To estimate gene flow between host races, we calculated the effective contributions of individuals from host $\beta$ to the gamete pool formed on host $\alpha$ (number of gametes produced by individuals from host $\beta$ over the total number of gametes produced on host $\alpha$). Mean interhost linkage disequilibrium among fitness loci also was calculated at each generation (D’) (36).

RESULTS

Deterministic model based on genotype frequencies: cases of one locus and two loci. We first investigated the possibility of sympatric divergence using model 1 (one type of locus). Regarding the one-locus case, the contribution of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ decreased slowly with increasing intensity of selection (Fig. 1). Gene flow between the two host races could be reduced significantly, but only with huge coefficients of selection: for example, $s$ needed to be close to 200 for gene flow between the two host races to be reduced below 1%.

In the case of two loci interacting multiplicatively (equation 1 with $\varepsilon = 1$), again using model 1, the contribution of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ decreased much more rapidly with increasing intensity of selection (Fig. 1). When the loci interacted multiplicatively, they could combine their effects to promote divergence between host races.

We then compared these conditions of sympatric divergence obtained with a life cycle without gamete dispersal with the ones required by life cycles with some, but restricted, gamete dispersal
or with unrestricted gamete dispersal. We investigated the effect of introducing a rate, $m$, of between-host gamete dispersal just before mating, with model 1 and two loci interacting multiplicatively (Fig. 2). As expected, when gametes dispersed at random between hosts ($m = 0.5$), there was unrestricted gene flow between individuals from the two host plants, even for huge selective coefficients. Lower rates of gamete dispersal prevented complete genetic isolation but allowed a certain degree of genetic isolation between host races.

When the two loci interacted additively (equation 2 with $\varepsilon = 1$), again using model 1 with no gamete dispersal ($m = 0$), the contribution of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ decreased much less rapidly with increasing intensity of selection than in the case with multiplicative interactions. To obtain a level of gene flow below 1%, the coefficient of selection had to be as high as $s = 5,000$ (data not shown). Thus, gene flow decreased even less rapidly with increasing intensity of selection on each locus when two loci interacted additively than when only one locus was considered. This is because, with one locus, only extreme genotypes exist whereas, with two loci, intermediate genotypes are produced on each host, which have intermediate fitness when loci interact additively. Therefore, these intermediate genotypes were able to produce gametes on both hosts, impeding reduction of gene flow between hosts. Thus, with additive interactions among loci, an increase of the number of loci did not facilitate divergence between host races and even led to far higher values of selection coefficient required per locus to reduce gene flow.

We then investigated whether divergence also was possible using model 2 (two types of loci). With one locus of each type ($2L = 2$) interacting multiplicatively (equation 3), these loci reached a stable level of linkage disequilibrium, just as in model 1 (data not shown). The contribution of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ depended here on both fitness parameters $s$ and $c$ (Fig. 3). Both $s$ and $c$ had to be high to significantly reduce gene flow (for instance, $c$ had to be higher than 0.8 and $s$ higher than 32 to reduce gene flow below 1%). When the two loci interacted additively using model 2 (equation 4 with $\varepsilon = 1$), gene flow remained above 30%, even with very high selective coefficients (data not shown). As expected, divergence between host races was more difficult using fitness loci having an allele with the same fitness in both hosts (model 2) than using fitness loci having one allele beneficial in one host and the other allele beneficial in the other host (model 1), but it was still possible, provided multiplicative interaction among loci and high values of selection coefficients.

**Stochastic model based on haplotypes: two and more loci.**

We first checked that similar results were obtained with the stochastic versions of the models as with the deterministic versions for the two-locus case. For instance, in the case of model 1 with two loci and without epistasis, the curves of the contribution of individuals from host $\beta$ to the gamete pool formed on host $\alpha$ obtained using individual-based models were very similar to those obtained with deterministic models (curves $L = 2$, Figs. 1 and 4). The stochastic models allowed us to observe the dynamics of reduction in gene flow and of evolution of linkage disequilibrium: equilibrium values were reached within a few generations (less than six; data not shown).

We then considered the cases with more than two loci. With multiplicative interactions, we observed again a build-up of interhost linkage disequilibria among loci (data not shown). Thus, with more fitness loci, lower intensities of selection at each locus were required to stop gene flow between the two host races (Figs. 4 and 5). With four or five loci, the curve already decreased very rapidly and even relatively low selective coefficients ($s = 1$) were sufficient to completely stop gene flow. Thus, the stochastic models showed that a few specialization genes, each having a small effect but acting multiplicatively, can be sufficient to promote host race formation and even speciation in sympatry. Here again, we checked that introducing unrestricted gamete dispersal prevented genetic isolation when there is no specific gene for active assortative mating, even with huge selective coefficients (Fig. 4, dotted lines). Also, a restricted gamete dispersal was allowed here to reduce gene flow significantly, but not completely.

![Fig. 2](https://example.com/fig2.png)

**Fig. 2.** Contribution at equilibrium of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ plotted against the coefficient of selection of adaptive alleles $(1 + s)$ with model 1 (deterministic version) for two loci ($L = 2$) acting multiplicatively (equation 1, $\varepsilon = 1$) for different values of gamete dispersal ($m = 0.5, 0.1, 0.01$, and 0.001).

![Fig. 1](https://example.com/fig1.png)

**Fig. 1.** Contribution at equilibrium of parasites coming from host $\beta$ to the gamete pool formed on host $\alpha$ plotted against the coefficient of selection of adaptive alleles $(1 + s)$ with model 1 (deterministic version) for one locus ($L = 1$, bold line) or two loci acting multiplicatively ($L = 2$, dashed line); equation 1, $\varepsilon = 1$. The solid line represents interhost linkage disequilibrium for the case of two loci ($D'$).
The effect of epistasis (\(\varepsilon\)) then was investigated (Fig. 6). Because similar results were obtained with both types of models, only results from model 1 are presented here. With a few loci (\(L = 3\)), gene flow decreased more rapidly with increasing intensity of selection when epistasis was synergistic (\(\varepsilon > 1\)) than without epistasis (\(\varepsilon = 1\)) and much less rapidly when epistasis was antagonistic (\(\varepsilon < 1\)). Increasing the number of loci reduced the impact of synergistic epistasis; because the curve obtained without epistasis was already maximally leptokurtotic with four or five loci (Fig. 4), increasing epistasis did not yield much reduction of gene flow. With five loci, for instance, the level of gene flow already was almost null for \(s = 3\) (Fig. 4); therefore, adding synergistic epistasis could not reduce the level of gene flow much more.

**DISCUSSION**

**Sympatric host race formation and speciation.** Our results show that gene flow between two populations of a parasite adapting to two hosts can be reduced in sympatry as a direct consequence of a strong specialization, without any active reproductive isolation mechanism (active host choice or active assortative mating), but only for organisms with a life cycle with restricted dispersal between development on the host and mating. This is because the low gamete dispersal induces a certain kind of pleiotropy between ecological specialization and assortative mating: the probability of mating between two individuals depends directly on their ability to infect the same host. This form of pleiotropy had not been investigated previously and may be widespread in pathogens. In previous models using pleiotropy, the probability of mating between two individuals depended on the difference in their phenotypic value (46) whereas, in our model, the probability of mating between two given genotypes depended on the product of their fitness. This difference is crucial because, when the probability of mating between two individuals depends on the difference in their phenotypic value, two individuals unfit on either host (i.e., with intermediate genotypes) can mate together. In contrast, when the probability of mating between two given genotypes depends on the product of their fitness, only the fittest individuals (i.e., with extreme genotypes) can produce offspring.

Moreover, despite some apparent similarities, our model is even more different from the previous ones using a classical life cycle with mating restricted within hosts: in these cases, sympatric divergence requires a gene for active host choice that has to come in linkage disequilibrium with fitness genes (27). The evolution of such linkage disequilibrium is often difficult because genes for

---

**Fig. 3.** Contribution at equilibrium of parasites coming from host \(\beta\) to the gamete pool formed on host \(\alpha\) plotted against the coefficient of selection of adaptive alleles (\(1 + s\)) with model 2 (deterministic version) and one locus of each type (\(2L = 2\)) acting multiplicatively (equation 3, \(\varepsilon = 1\)) for different values of the coefficient of selection of nonadaptive alleles (\(c = 0.3, 0.5, \) and 0.8).

**Fig. 4.** Contribution at equilibrium of parasites coming from host \(\beta\) to the gamete pool formed on host \(\alpha\) plotted against intensity of selection per locus (\(1 + s\)) with model 1 (stochastic version) and loci acting multiplicatively (equation 1, \(\varepsilon = 1\)) for different numbers of locus (\(2L = 2, 4, 6, 10, \) and 20). The parameter \(c\) was set to 0.5.

**Fig. 5.** Contribution of parasites coming from host \(\beta\) to the gamete pool formed on host \(\alpha\) plotted against intensity of selection (\(1 + s\)) with model 2 (stochastic version) and loci acting multiplicatively (equation 1, \(\varepsilon = 1\)) for different values of the coefficient of selection of nonadaptive alleles (\(c = 0.3, 0.5, \) and 0.8).

**Fig. 6.** Contribution at equilibrium of parasites coming from host \(\beta\) to the gamete pool formed on host \(\alpha\) plotted against intensity of selection (\(1 + s\)) with model 1 (stochastic version) and three loci (\(L = 3\)) for different values of epistasis (equation 1, \(\varepsilon = 0.5, 1, \) and 1.5).
host choice (i.e., genes for active recognition of a mating partner) are not directly subjected to selection. In our model, only host-based fitness genes were simulated that were selected for directly.

Therefore, in pathogens with restricted dispersal between development on the host and mating, the existence of genetic variability for host exploitability should lead directly to host race formation. Thus, the process simulated here may explain the widespread occurrence of host races and specialized sibling species in pathogens (5,6,18,24,33,45,47,48,53). Indeed, many fungal species may have some restriction in gamete dispersal compared with that of asexual spores (1). A restricted but not null gamete dispersal may explain the existence of fungal populations with only slight genetic differentiation on different crops (7,8). Provided very high selective coefficients and trade-offs, even speciation could be achieved in our model. The results of the stochastic simulations showed that, in finite populations, effective contributions of parasites from one host to the gamete pool on another could reach zero; for sufficiently high selective coefficients, gene flow between the two host races stopped (across the whole genome), so that they could be considered as independent evolving units. This occurred when individuals dispersing from each host were sufficiently specialized to be virtually unable to develop, and thus mate, on the other host, compared with individuals that dispersed from the same host. Therefore, specialization had to be much stronger than in previous models of sympatric speciation, where values of selective coefficients s required for speciation usually were below 1. One good example of this process may be the ascomycete V. inaequalis (33): gametes are produced within the leaves and, therefore, fit our model of lack of gamete dispersal. Formae speciales recently have been described in V. inaequalis that have no prezygotic isolation (i.e., are completely interfertile in vitro) but are highly specialized on two different plants, apple and pyracantha. Molecular markers showed that they do not exchange genes at all in sympatry (33). Of existing models, only the one proposed here can explain how such a pattern of reproductive isolation can arise, though the idea is similar to the “natural selection against immigrants from divergent habitats” recently proposed to occur in animals from a literature survey (40).

Thus, in taxa with no dispersal between selection and mating, strong specialization alone is able to create genetic isolation (restriction on gene exchange, even for neutral genes). This contrasts with taxa with classical life cycles, for which even similarly huge selective coefficients are incapable of impeding neutral gene flow, as shown in our simulations with gamete dispersal. When gametes disperse at random between the two hosts (m = 0.5), as is the case in classical life cycles, no restriction in neutral gene flow was possible at all, even for huge selective coefficients. High selective coefficients maintain the adaptive alleles in each habitat; however, because individuals disperse before mating, alleles at neutral loci still can be exchanged between individuals from the two hosts, unless there are specific genes for assortative mating. Therefore, with classical life cycles, strong selection allows maintenance of polymorphism in the population, as shown by previous models (35), but does not allow for divergence of host races. True divergence can occur only with evolution of active assortative mating or active host choice if mating is restricted within hosts. This may be the case of the “good biological” sibling species described in some fungal taxa, with unrestricted gamete dispersal and having strong prezygotic isolation among sibling species; for instance, in the genus Armillaria (2).

Here, we simulated strong specialization either by a few loci with high selective coefficients or several fitness loci acting multiplicatively. In the latter case, strong specialization could be achieved because the lack of dispersal between selection and mating created extreme interhost linkage disequilibria between fitness loci, despite initial possibility of gene flow. Selection increased the number of adaptive alleles in each host just before crosses and, because there was no host shift between selection and mating, it led to high frequencies of favorable allelic combinations on each host and, thus, interhost linkage disequilibria. At each generation, the fittest combinations of alleles reproduce most in each host, creating, after crosses, even better combinations of alleles. Over generations, ever fitter allelic combinations are created in each host, and the fitness discrepancy increases between the pools of individuals coming from the two habitats, increasing in turn linkage disequilibria and differences in contributions to the gamete pools. Thus, the possibility to create linkage disequilibria also is a specificity provided by the particular life cycle without gamete dispersal; in models of sympatric speciation using the life cycle of free-living animals, increasing the number of loci usually restricts the possibility of divergence (22).

Our model differs from allopatric models of speciation in that selection and not geographical isolation impeded mating between individuals: at each generation, offspring dispersed freely between hosts, but only sufficiently adapted individuals could develop and, thus, mate within a given host. Therefore, divergence was truly sympatric in our model but yielded patterns resembling those produced by allopatric speciation in some aspects. Speciation is caused by a complete cessation of gene flow without active assortative mating, just as in allopatric speciation; therefore, the conditions creating this genetic isolation must be maintained a certain time to allow the incipient species to diverge before the new species can be fully recognized. Once gene flow is strongly restricted by specialization, genomic incompatibilities will accumulate, leading to intrinsic reproductive isolation and then to complete reproductive isolation, just as in allopatric divergence (42).

**Plausibility of the different parameters and hypotheses.** In fungi, the hypotheses of model 1, where, at each locus, one allele gives an advantage on one host and the other allele on the other host, are valid for genes coding for some toxins. It will be the case for some toxins that play an important role, for instance in pathogenicity of the fungus Fusarium graminearum (38). Alleles of the same cluster of genes can induce the production of different types of toxins: Nivalenol is produced when the TRI 7 and TRI 13 genes are functional, whereas deoxynivalenol is produced when these genes are nonfunctional (34). These different toxins may give advantages on different host plants (8,43), as assumed in model 1, although this role is controversial. Indeed, the different Fusarium chemotypes are frequently found in association on the same host plants (26) and the trichothecene chemotype polymorphism is transspecific and ancestral, indicating balancing selection operating within Fusarium spp. for a long time (54).

In fungi, most of the genes acting on pathogenicity are more consistent with model 2, with one allele beneficial on one host and the second allele having the same effect on all hosts. For instance, such hypotheses apply well to most toxin production genes (55) and avirulence genes (25). Avirulence genes, very common in pathogenic fungi, produce an avirulence product that is recognized by resistant plants, inducing a hypersensitivity reaction that impedes infection. The alternative allele does not produce any avirulence product and, therefore, has the same fitness on both types of plant. Similarly, most toxin production genes usually have an allele that gives an advantage in one host (toxin is required for successful infection) and an alternative allele with no function and, hence, the same fitness on both hosts. In fungal Alternaria spp., for instance, a host-specific toxin is required for successful infection of a given host, the different toxins being produced by different sets of genes, and polymorphism for toxin production exists in natural isolates (28).

The values of the parameters allowing sympatric divergence in our model are plausible in fungi. The selective coefficients required for divergence are much higher than in most models of speciation, but such values are not unrealistic in phytopathogenic fungi. Numerous examples are given in the literature of among-strains differences in infection ability that are consistent with the
high selective coefficients required in our model. For instance, avirulence alleles can completely impede infection (25) and toxin production alleles can provide a huge advantage for infection on a given plant or even be absolutely required (38). This corresponds to virtually infinite fitness differentials. Moreover, when a few genes act multiplicatively, the difference in fitness values of the alleles need not be too high, and experimental studies have shown that several genes can be involved in pathogenicity. For instance, multiple loci coding for mycotoxins with strong effects on pathogenicity or infection ability are found in the *Fusarium* spp. complex (3,39). The production of each of these different toxins is controlled by a specific gene cluster comprising many genes involved in toxin biosynthesis pathways (3,39). These genes act in a multiplicative manner because each step is dependent on the previous one. Because they are organized in clusters, they probably are most often inherited as units. However, this organization in clusters may be have arisen secondarily and there may still be some occasional recombination (9). Thus, the conditions required for divergence and even subsequent speciation in our model (either low number of loci with strong selection or many loci with weaker selection interacting multiplicatively) are realistic in light of the existing literature on fungi.

**Consequences for species criteria.** The questions of what species are (species concept) and how one can recognize them (species criteria) have long been debated and have often been confounded (13). It has been argued that the numerous species concepts (e.g., biological species concept [BSC], phylogenetic species concept [PSC], and morphological species concept [MSC]) are, in fact, variants of the same species concept (evolutionary independent lineages) (13). Indeed, the main problem seems less to find a universal species definition than to find operational species criteria (i.e., to identify species in the field). There exist numerous criteria to distinguish species. The most famous one is active reproductive isolation, on which the BSC is based.

Previous models of sympatric speciation supported this criterion because they showed that speciation required the evolution of active intrinsic reproductive isolation. In contrast, the mechanism of sympatric divergence that we propose here does not assume the evolution of a specific gene responsible for reproductive isolation. This corresponds well to fungal host-specialized cryptic "species" (47-49) which often are able to interbreed in the laboratory but, nevertheless, are genetically isolated even in sympatry in nature. Thus, our model may explain why criteria using molecular phylogenies (referring to the PSC) appear more useful in fungi (15,48,49) and, more generally, in parasites (12) than morphological (referring to the MSC) or intersterility criteria (referring to the BSC). With our mechanism of speciation, we expect that the ancestral ability to interbreed will be retained a certain time after effective genetic isolation between two incipient species arises, as is also expected in allopatric speciation. Therefore, our model will have important consequences for the choice of species criteria in parasites.

**ACKNOWLEDGMENTS**

T. Giraud and L. M. M. A. Villarié contributed equally to the work in this paper and should be considered as sharing the position of first authorship. This work was supported by a grant from the French Ministère de la Recherche (“ACI Jeunes Chercheurs 2003”). We thank J. A. Shykoff, Y. Brygou, E. Fournier, and J. Enjalbert for their invaluable help and stimulating discussions on the topic; and T. Kawecki, A. Kondrashov, M. Kirkpatrick, S. Gavrilets, C. Combes, S. Gandon, L. Keller, J. Koella, B. Genton, J. Correll, and three anonymous referees for helpful comments on previous versions of the manuscript.

**LITERATURE CITED**


35. Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. Am. Nat. 87:331-333.