Evolution, 57(8), 2003, pp. 1735–1746

MULTILOCUS MODELS OF SYMPATRIC SPECIATION: BUSH VERSUS RICE VERSUS FELSENSTEIN

JAMES D. FRY
Department of Biology, University of Rochester, Rochester, New York 14627-0211
E-mail: jfry@mail.rochester.edu

Abstract.—In populations of phytophagous insects that use the host plant as a rendezvous for mating, divergence in host preference could lead to sympatric speciation. Speciation requires the elimination of ''generalist'' genotypes, that is, those with intermediate host preference. This could occur because such genotypes have an inherent fitness disadvantage, or because preference alleles become associated with alleles that are oppositely selected on the two hosts. Although the former mechanism has been shown to be plausible, the latter mechanism has not been studied in detail. I consider a multilocus model (the ''Bush model'') in which one set of biallelic loci affects host preference, and a second set affects viability on the hosts once chosen. Alleles that increase viability on one host decrease viability on the other, and all loci are assumed to be unlinked. With moderately strong selection on the viability loci, preference alleles rapidly become associated with viability alleles, and the population splits into two reproductively isolated host specialist populations. The conditions for speciation to occur in this model, as measured by the strength of selection required, are somewhat more stringent than in a model in which preference and viability are controlled by the same loci (one-trait model). In contrast, the conditions are much less stringent than in a model in which speciation requires buildup of associations between viability loci and loci controlling a host-independent assortative mating trait (canonical two-trait model). Moreover, in the one-trait model, and to a lesser extent the Bush model, the strength of selection needed to initiate speciation is only slightly greater than that needed to complete it. This indicates that documenting instances of sympatric species that are reproductively isolated only by host or habitat preference would provide evidence for the plausibility of sympatric speciation in nature.

Key words.—Ecological speciation, habitat isolation, host race, hypergeometric phenotypic model, pea aphids, phytophagous insects.

Received November 12, 2002. Accepted February 23, 2003.

Whether speciation can occur without geographic isolation has been debated since the 19th century. Although Darwin was a supporter of sympatric speciation (Darwin 1859; Kondrashov et al. 1998), it had fallen out of favor among many, if not most, evolutionary biologists by 1980. Two factors contributed to this. Convincing empirical evidence for sympatric speciation was lacking (e.g., Futuyma and Mayer 1980), and there was a widespread view that sympatric speciation requires extraordinarily stringent conditions (Mayr 1963; cf. Felsenstein 1981). In the last 20 years, however, the tide has shifted somewhat. On the empirical side, there are now two reasonably well-documented cases in which sympatric speciation, or at least incipient speciation, seem probable (Schliewen et al. 1994; Feder et al. 1988, 1994; Filchak et al. 2000). On the theoretical side, models of sympatric speciation have become more sophisticated, incorporating multilocus genetics (e.g., Kondrashov 1983a,b, 1986; Rice 1984; Doebeli 1996; Kondrashov et al. 1998; Kondrashov and Kondrashov 1999; Dieckmann and Doebeli 1999), and a greater degree of ecological realism (e.g., Diehl and Bush 1989; Doebeli and Dieckmann 2000) than early models (e.g., Maynard Smith 1966; Felsenstein 1981).

An appealing class of sympatric speciation models is that in which disruptive natural selection on a polygenic character causes a population to split into two reproductively isolated populations. (Although there is debate about the most useful definition of the term ''species'' [e.g., Mallet 1995; Harrison 1998], no matter which definition is adopted, some degree of reproductive isolation is a necessary condition for two *sympatric* populations to coexist as separate species). The disruptive selection can be a fixed feature of the environment (e.g., Kondrashov 1983a,b, 1986; Rice 1984; Kondrashov et al. 1998; Kondrashov and Kondrashov 1999), or the result of intraspecific competition or other ecological interactions (Doebeli 1996; Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2000). Reproductive isolation in these models builds up in either of two basic ways (for a review, see Kirkpatrick and Ravigné 2002). In one-trait or single-variation models, the trait under disruptive selection simultaneously serves as a basis for assortative mating; therefore, genotypes with intermediate values of the mating trait are directly selected against. Such models, which have been forcefully advocated by Rice (1984; Rice and Hostert 1993), result in speciation comparatively easily, because the combination of direct selection and assortative mating is fairly effective at eliminating the intermediate genotypes. The feasibility of one-trait models has been confirmed in several laboratory experiments on speciation (reviewed in Rice and Hostert 1993; Kirkpatrick and Ravigné 2002). In contrast, in two-trait or double-variation models the trait under selection does not influence mating; instead, assortative mating occurs for a second trait (the AM trait), which is usually assumed to be neutral. Nonetheless, if alleles affecting the AM trait develop chance associations with alleles affecting the selected trait, genotypes with intermediate values of the AM trait will also tend to have intermediate values of the disruptively selected trait, resulting in indirect disruptive selection on the AM trait. When the AM trait is neutral, however, strong disruptive selection on the other trait is needed for speciation, a point made forcefully by Felsenstein (1981). The conditions for speciation are even more stringent in three-trait models, in which assortative mating depends on two sex-specific traits: a preference trait in females and a signaling trait in males (Kondrashov and Kondrashov 1999). Whether the type of

indirect selection envisioned in these two- and three-trait models can lead to sympatric divergence has been tested in the laboratory by applying strong disruptive selection to a trait not believed to be associated with mating, for example, bristle number in *Drosophila*. With one exception (Thoday and Gibson 1962), many such experiments have failed to produce significant assortative mating between the selected extremes (reviewed in Thoday and Gibson 1970; Scharloo 1970; Rice and Hostert 1993).

Based on theoretical and empirical work to date, therefore, the one-trait model is the most plausible model of sympatric speciation. This model has a serious limitation, however: although disruptive selection is common in natural populations (Endler 1986; Kingsolver et al. 2001), the requirement that the same genes underlie the disruptively selected trait and assortative mating is restrictive. It is probably safe to say that the average trait under disruptive selection does not automatically generate positive assortative mating (cf. Diehl and Bush 1989). Modifier alleles that cause AM on the basis of the trait will theoretically be favored (Kondrashov 1986; Doebeli 1996; Dieckmann and Doebeli 1999), but the failure of the numerous laboratory tests of indirect selection suggests that such modifiers are uncommon. Therefore, any broadly applicable model that lacks the requirement of direct disruptive selection on the AM trait, but results in speciation more readily than typical two-trait models, would be of great interest.

A candidate for such a model has been proposed verbally by Bush (1975) and analyzed to a limited extent by Diehl and Bush (1989; also see Kawecki 1996, 1997). The model, which I call the ''Bush model,'' applies to any population that uses two discrete habitats or hosts, and in which mating takes place within the habitats or on the hosts. Its most obvious application is to phytophagous insects, many of which mate exclusively or almost exclusively on their host plants (Bush 1975). The model assumes that there are one or more loci at which opposite alleles are favored on the different hosts, a situation sometimes termed ''diversifying selection.'' These loci do not affect host preference or mating behavior. A second locus, or set of loci, determine host preference, but do not affect fitness on a host once it is chosen. Under these assumptions, alleles favored on a given host are expected to become associated with alleles conferring preference for the host, a prediction confirmed by Diehl and Bush (1989). Once this linkage disequilibrium builds up, genotypes with intermediate host preference will be at a fitness disadvantage compared to genotypes with strong preference for one host or the other (cf. Kawecki 1996, 1997). If all genotypes with intermediate host preference are eliminated, speciation will be complete. The Bush model differs from one-trait models in that genotypes with intermediate values of the AM trait, host preference, are not selected against until linkage disequilibrium builds up between the preference loci and the fitness-determining loci. The model differs from ''canonical'' two-trait models (e.g., Maynard Smith 1966; Dickinson and Antonovics 1973; Felsenstein 1981; Kondrashov and Kondrashov 1999) in that the direction of linkage disequilibrium between host-preference loci and fitness loci is not arbitrary: even with no initial linkage disequilibrium, alleles conferring higher fitness on a host become associated with alleles conferring preference for that host (Diehl and Bush 1989). This is because, by affecting whether an individual chooses a host to which it is relatively well adapted, host preference interacts with the trait under diversifying selection to determine fitness.

An attractive feature of the Bush model is that its basic assumptions are likely to be met in many phytophagous insect populations. Mating on the host is widespread in phytophagous insects (Bush 1975; Berlocher and Feder 2002), as is genetic variation for host preference (Jaenike and Holt 1991). Furthermore, different host species are known to select for different morphological traits in phytophagous insect populations (Clarke et al. 1963; Moran 1986; Carroll and Boyd 1992; Carroll et al. 1997), and almost certainly select for different behavioral (e.g., Messina 1991; Caillaud and Via 2000) and physiological (e.g., Karowe 1990; reanalyzed in Fry 1992) traits in some instances.

In spite of the potential widespread applicability of the Bush model, previous analyses have given little information on how readily sympatric speciation occurs under its assumptions compared to one-trait and typical two-trait models. This is because the main published analyses of the model have assumed that host preference is controlled by a single locus (Diehl and Bush 1989; Kawecki 1996, 1997). This is a severe limitation, for two reasons. First, with only a single host-preference locus, the amount of reproductive isolation that can build up is limited (cf. Kondrashov 1986); in fact, complete reproductive isolation was not a possible outcome of the Diehl and Bush (1989) model (see Discussion). Second, without recombination to regenerate genotypes with intermediate host preference, there is no factor opposing speciation. As a result, in models with only a single preference locus, there is no threshold strength of selection under which progress toward speciation does not occur. The Bush model is briefly considered by Kondrashov (1983b, 1986), but is not treated in detail. The model of Johnson et al. (1996) has two host-preference loci, but also has an additional, host-independent assortative mating trait, and does not allow speciation to occur by host isolation alone.

Here I compare the feasibility of the Bush model, the one-trait model, and the canonical two-trait model in a multilocus framework. Specifically, I ask how the minimum strength of disruptive or diversifying selection necessary for speciation differs between the three models. For each model, I also investigate to what extent the strength of selection necessary for sympatric speciation to occur in an initially panmictic population is greater than that necessary to prevent two differentiated species from fusing after a small amount of hybridization. If there is little difference between the two sets of conditions in a particular model, then the observation of two sympatric species that are reproductively isolated by the mechanism hypothesized in the model would suggest that the mechanism could have caused the initial divergence in sympatry. In addition, I examine the effects of different life cycles found in phytophagous insects on the likelihood of speciation by host isolation.

In the next section, I study the behavior of a version of the Bush model with four preference loci and two fitness loci. Subsequently, I generalize the model to different numbers of loci. Finally, I develop parallel versions of the one-trait and

canonical two-trait models for comparison with the Bush model.

An Exploratory Version of the Bush Model

I consider a population of phytophagous insects that uses two equally abundant host species, A and B. Mating occurs on the host after hosts are chosen. Initially, however, insects do not discriminate between the hosts, and have equal fitness on them.

In this section, I examine the fate of alleles that alter host preference and host-specific fitness when these are introduced at low frequency. For simplicity, I assume haploidy. Host preference is controlled by four host-preference loci, each with two alleles, + and −. An individual with i + alleles chooses host A with probability $i/4$, and chooses host B with probability $1 - i/4$. Initially, the population is fixed for + alleles at two of the preference loci, and for − alleles at the other two. Fitness on a host is unaffected by the preference loci. Instead, two additional loci determine survival on a host once it is chosen. Each viability locus has a + and − allele; + alleles confer higher survival on host A at the expense of ability to survive on host B, whereas − alleles have the opposite effect. Specifically, an individual with i + alleles survives on host A with probability $(1 - s)^{(2-i)}$, but survives on host B with probability $(1 - s)^i$. Initially, the population is fixed for the + allele at one of the viability loci and for the − allele at the other. All loci are assumed to be unlinked. Selection is assumed to be ''soft,'' with each host contributing one-half of the individuals to the dispersing pool. Although soft selection can result in unrealistically strong frequency-dependent selection on host preference (e.g., Rausher and Englander 1987), this happens only when one host is used by a much smaller percent of the population than the other, a situation not considered here.

Speciation occurs in this model if the population comes to consist of only two genotypes at the preference loci, one with four + alleles and the other with four − alleles; these host specialist genotypes will not encounter each other and thus will be reproductively isolated. Speciation thus requires buildup of complete linkage disequilibrium among all four preference loci.

It could be argued that it would be more relevant to assume that the initial population accepts only host A, and then examine the conditions under which a second species that accepts only B is produced. It is difficult to imagine, however, how a population could evolve from accepting host A and rejecting B, to rejecting A and accepting B, without going through a period in which it accepted both hosts. Even if this period were a few thousand generations, it would be practically instantaneous on a geological time scale. My model bypasses the question of how the initial host range expansion occurred, because it is the subsequent step, the splitting of the generalist population into two host-specific populations, in which the evolution of reproductive isolation takes place.

I consider two alternative life cycles, differing in whether selection occurs before dispersal and host choice, or after. With selection before host choice, the life cycle is: newborns → viability selection on host → dispersal and host choice → random mating on host → zygotes → meiosis → newborns.

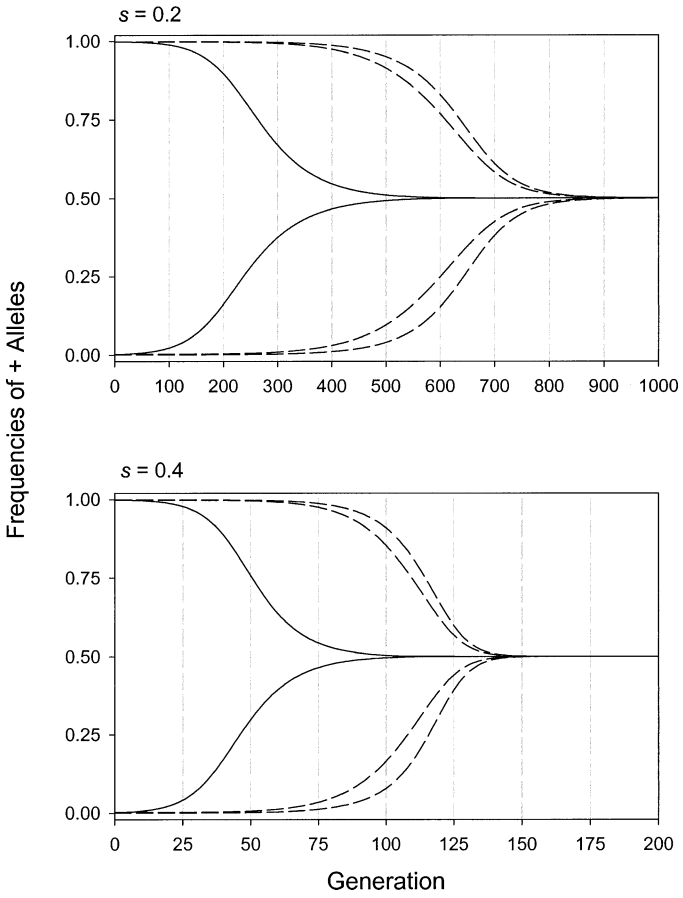

FIG. 1. Changes in frequencies of + alleles in the exploratory version of the Bush model described in the text. Solid and dashed lines give frequencies at the viability and preference loci, respectively. Initial + viability allele frequencies were 0.002 and 0.999. Initial + preference allele frequencies were 0.001, 0.003, 0.998, and 0.999. All rare alleles initially existed only in conjunction with the common alleles at other loci.

This life cycle resembles that in many Diptera and Lepidoptera, in which adults choose the host for the larvae, but do not feed on the host. With selection after host choice, the life cycle is: newborns → dispersal and host choice → selection → mating → zygotes → meiosis → newborns. Life cycles like this do not appear to be common in insects (although that of scale insects, in which first instar larvae disperse and subsequent stages, except for adult males, are wingless and legless, comes close). Nonetheless, in many phytophagous insects in which adults are the dispersing stage (e.g., aphids, many beetles), both adults and immatures feed on the host, so selection occurs both before and after host choice. The model with selection after host choice allows the effectiveness of postdispersal selection in promoting sympatric speciation to be studied in isolation.

To determine the consequences of introducing each of the alternative alleles at low frequency, a Mathematica program (Wolfram 1996; available from J. Fry on request) was written to perform deterministic iterations of the complete set of 64 genotype frequencies, assuming an effectively infinite population. Figure 1 shows changes in the six allele frequencies over time for $s = 0.2$ and $s = 0.4$ (see the figure legend for

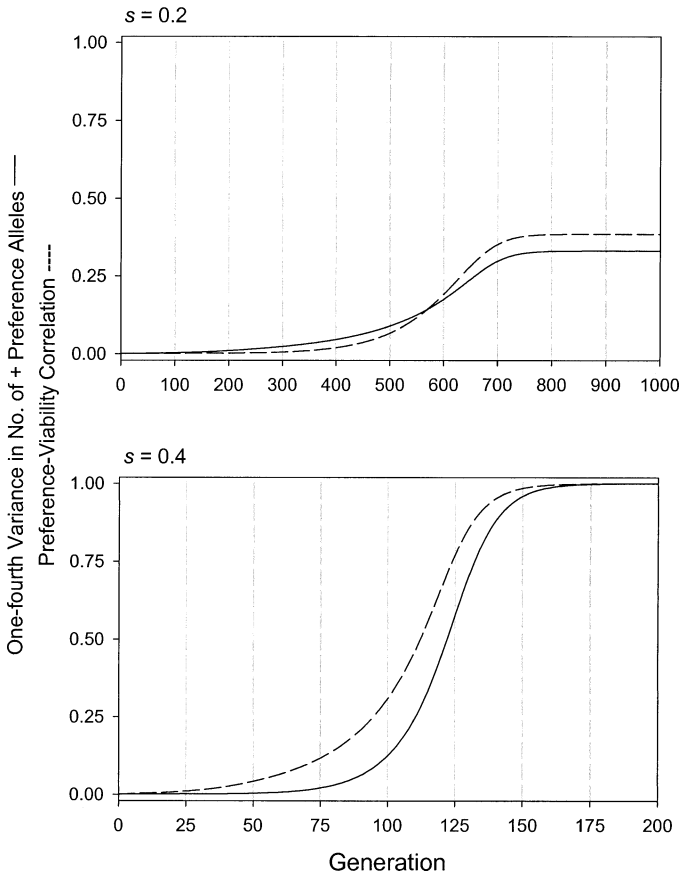

FIG. 2. One-fourth the variance in the number of + preference alleles (solid lines), and the correlation between the number of + preference and + viability alleles (dashed lines), in the exploratory model.

starting genotype frequencies). These results are for the life cycle with selection after host choice; the alternative life cycle gave qualitatively similar results (not shown). With either selection coefficient, all six rare alleles increased in frequency, until all allele frequencies reached 0.5. Viability loci converged to 0.5 more quickly than preference loci. Allele frequencies changed more slowly with lower s, but as long as s was greater than zero, convergence of allele frequencies to 0.5 eventually occurred (results not shown). In contrast, allele frequencies remained essentially unchanged when s was 0. The latter result confirms that the preference alleles do not invade because of an inherent rare-allele advantage. Instead, polymorphism at the preference loci is apparently maintained because + (−) preference alleles become associated with + (−) viability alleles, causing selection on the preference loci to qualitatively resemble selection on the viability loci.

Progress toward speciation was monitored by calculating the variance in the number of + preference alleles among newborns, and the correlation between the number of + preference alleles and + viability alleles (Fig. 2). When speciation is complete (no gene flow between hosts), half of the individuals in the population have zero + preference alleles, whereas the other half have four; the variance in the number of preference alleles is therefore $(0.5)(0.5)4^2 = 4$. (Once allele frequencies have reached 0.5, this variance is a linearly increasing function of the average linkage disequilibrium among the six pairs of preference loci. With no linkage disequilibrium, the variance would be 1). Similarly, when speciation is complete, all individuals with + (−) preference alleles have only + (−) viability alleles; the correlation between the number of + preference and + viability alleles is therefore 1.

Figure 2 shows that speciation occurs for $s = 0.4$, but not for $s = 0.2$. In the former case, although all allele frequencies were between 0.4 and 0.6 by generation 126 (Fig. 1), the preference variance and the preference-viability correlation continued to increase for about another 25 generations (Fig. 2). During this phase, evolution of the population occurred primarily by buildup of linkage disequilibrium. In contrast, with $s = 0.2$, once all allele frequencies were between 0.4 and 0.6 (generation 712; Fig. 1), there was little further change in either the preference variance or the preference-viability correlation, both of which had leveled off at about one-third of their maximal values (Fig. 2). At this point, gene flow between hosts remained high, with about 30% of those choosing each host having originated from the other.

By trial and error, the threshold for speciation was found to lie between $s = 0.265$ and $s = 0.270$. Below this threshold, selection is too weak to overcome the homogenizing effect of recombination. For the alternative life cycle, with selection before host choice, stronger selection ($0.510 < s < 0.515$) was required for speciation. These thresholds were not sensitive to initial genotype frequencies, as long as linkage disequilibrium was initially slight or absent (see Appendix).

Extension to Different Numbers of Loci

The above results show that with four preference loci and two viability loci, sympatric speciation can occur by host isolation, although strong selection is required to overcome the selection-recombination antagonism (Rice and Hostert 1993). However, both preference and viability are likely to be affected by more than two to four loci (e.g., Jones 1998; Hawthorne and Via 2001). Therefore, it is important to consider how the conditions for speciation change as the number of loci is increased.

In a model with n preference loci and m viability loci, there are 2^{n+m} genotypes; thus following frequencies of all genotypes becomes impractical unless n and m are small. However, such a model has only $(n + 1)(m + 1)$ phenotypes defined by the number of + alleles of each type. If within phenotypes, each genotype is equally frequent (the equiprobable state), then phenotype frequencies alone give a sufficient description of the system. In this case, properties of the hypergeometric and binomial distributions can be used to write recursions for the phenotype frequencies (Shpak and Kondrashov 1999; Barton and Shpak 2000).

Models of disruptive selection with equal allelic effects, free recombination, and an element of frequency-dependent selection appear to evolve toward the equiprobable state as a general rule (Shpak and Kondrashov 1999; Barton and Shpak 2000). These properties are shared by most models of sympatric speciation, including the exploratory model considered above. As shown in the Appendix, convergence of

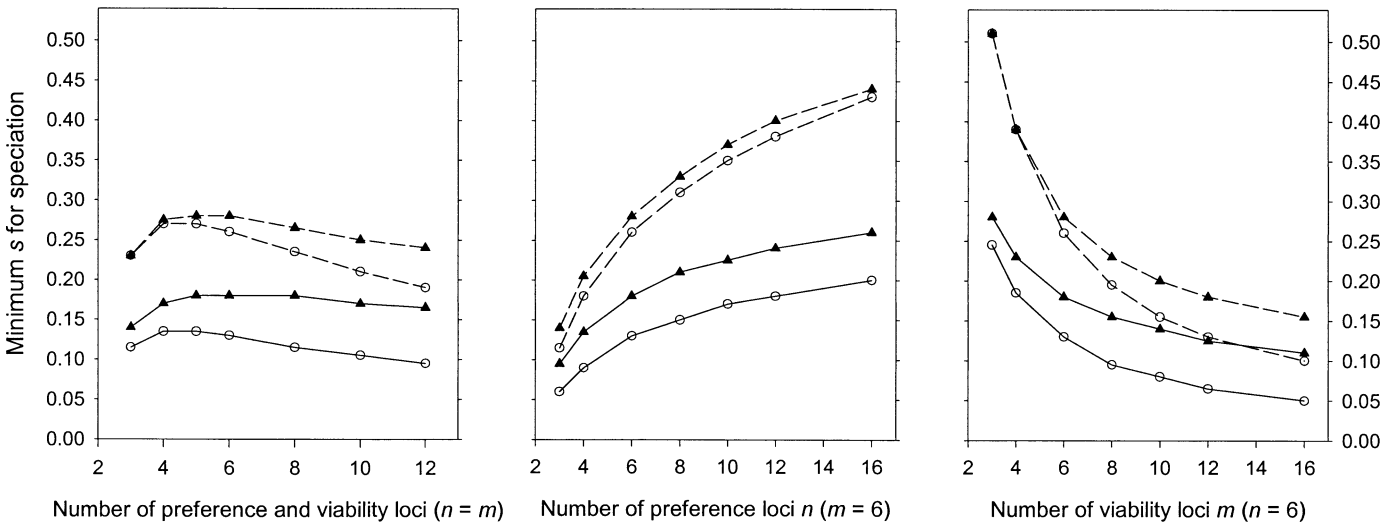

FIG. 3. Results of the Bush model for different combinations of n and m, in terms of the minimum s necessary for speciation. Solid lines, selection after host choice; dashed lines, selection before host choice. Triangles, minimum s for entire speciation process, starting with population in linkage equilibrium. Open circles, minimum s for speciation completion only, starting with a population consisting mostly of the two extreme genotypes (see text).

allele frequencies to 0.5 and achievement of the equiprobable state occurred from a wide range of starting genotype frequencies in this model. The same results were observed with $n = m = 3$ (see Appendix); there seems to be no reason why increasing the numbers of preference and viability loci would alter this result (Shpak and Kondrashov 1999; Barton and Shpak 2000). Therefore, the approach of following phenotypic frequencies (hereafter, the phenotypic method) can be used to determine the minimum s necessary for speciation for various numbers of preference and viability loci. In particular, the phenotypic method can be used to study the dynamics of phenotype frequencies in the second phase of evolution shown in Figures 1 and 2, in which allele frequencies are 0.5, and further evolution through changes in linkage disequilibrium either occurs or does not occur.

To generalize the model to n preference loci, I assumed that a genotype with i + preference alleles chooses host A with probability i/n, and chooses host B with probability $1 - i/n$. Similarly, with m viability loci, viability of an individual with i + viability alleles is $(1 - s)^{(m-i)}$ on host A, and remains $(1 - s)^i$ on host B. Given equiprobability of genotypes within phenotypes, the exact frequencies of phenotypes resulting from matings between any two phenotypic classes are given by equations 1–3 in Shpak and Kondrashov (1999). A Mathematica program (Wolfram 1996; available from J. Fry on request) incorporating these equations was written to iterate population phenotype frequencies, taking into account the selection and dispersal/host choice steps.

Progress to speciation was studied under two sets of starting conditions. To determine the minimum s necessary to both initiate and complete speciation, the starting population was assumed to have no linkage disequilibrium, with all allele frequencies equal to 0.5. In such a population, the two extreme phenotypes are rare. Phenotype frequencies were iterated until either (1) the two extreme genotypes had combined frequencies >0.999, in which case speciation was considered to have occurred; or (2) the sum of absolute values of the frequency changes from one generation to the next was less than 10^{-7}. In this case frequencies of the extreme genotypes were always $\ll 0.999$, so speciation was considered not to have occurred. The minimum s for speciation was found by varying s in increments of 0.005, unless $s > 0.3$, in which case increments of 0.01 were used.

The minimum s necessary to complete speciation in a population consisting mostly of the two extreme genotypes was also determined. In this case, the initial population was assumed to have the two extreme genotypes in frequencies of 0.495 each; the remaining 1% of the population had the phenotypic distribution corresponding to equal allele frequencies and absence of linkage disequilibrium. Phenotype frequencies were iterated until the direction of change was apparent (either elimination or increase in frequency of the intermediate phenotypes).

Both types of minimum s values are shown in Figure 3 for various values of n and m. The following conclusions emerge from these results:

(1) When selection occurs before host choice (dashed lines in the figures), the conditions for speciation are more stringent than when selection occurs after host choice (solid lines). In particular, minimum s values for speciation completion (open circles) are consistently about two-fold higher in the former life cycle than in the latter, whereas minimum s values for the entire speciation process (solid triangles) are somewhat less than two-fold higher.

(2) Although the conditions for speciation completion are usually less stringent than those for the entire speciation process, the differences are sometimes slight. In particular, the difference is small when selection acts before host choice, and with low numbers of viability loci (right-hand figure). The difference is not strongly affected by the number of preference loci (middle figure).

(3) Increasing the number of preference loci increases the minimum s for speciation (middle figure). This is not surprising, because by assumption, the effect of each preference

locus is inversely related to the number of loci. Increasing the number of preference loci therefore weakens selection on each one, and makes it more difficult for selection to build up the requisite linkage disequilibrium.

(4) Increasing the number of viability loci reduces the minimum *s* for speciation (right). This is also not surprising, because the net selection against intermediate phenotypes becomes stronger as the number of viability loci increases (but see below).

(5) If the number of preference and viability loci is assumed equal, the minimum *s* for speciation changes little as the number of loci is varied (left), presumably because the two above effects roughly cancel.

In general, the results in Figure 3 indicate that when selection acts after host choice, only moderately strong selection ($0.20 < s < 0.10$) is needed to cause sympatric speciation, unless the number of preference loci considerably exceeds the number of viability loci. When selection acts before host choice, stronger selection ($s > 0.20$) is often needed.

One way to put the minimum selection coefficients into perspective is to consider how low the fitness of ''generalists'' has to be for speciation to occur, where a generalist is defined as any genotype with equal or close to equal numbers of + and − viability alleles (depending on whether *m* is even or odd). For even *m*, the fitness of generalists is $(1 - s)^{m/2}$. For the combinations of *n* and *m* shown in Fig. 3, when selection acts after dispersal, speciation occurs when the fitness of generalists is 34–81% that of the specialist genotypes, those with all viability and preference alleles of one type. With selection before host choice, the corresponding range is 18–70%. Interestingly, increasing the number of viability loci lowers the critical fitness of generalists above which speciation cannot occur; by this measure, the conditions for speciation become more stringent as the number of viability loci is increased.

To put the last point in another way, for a given fitness difference between specialists and generalists, or between the two specialists on the host of one of them, increasing the number of viability loci makes speciation less likely. The likely explanation for this is that spreading the same fitness difference over more freely recombining loci weakens the selection on each one, and increases the effectiveness of recombination in breaking down linkage disequilibrium. The effect of increasing the number of viability loci while holding the fitness difference constant is therefore analogous to that of increasing the number of preference loci.

Comparison to the One-Trait and Canonical Two-Trait Model

To permit a formal comparison of the conditions necessary for speciation between the Bush, one-trait, and canonical two-trait models, I modified the assumptions of the Bush model to produce versions of the latter two models.

To produce a version of the one-trait (1T) model, host preference and viability were assumed to be controlled by the same *n* loci. At each locus, + alleles simultaneously increase viability on host A and preference for host A, whereas − alleles have the opposite effects. All other assumptions were unchanged. In this model, there is direct disruptive selection on host preference, the assortative mating trait; genotypes with all + or all − alleles automatically have the highest fitness. Although host preference is still the AM trait in this model, the model differs only in relatively minor details from one-trait models with other forms of assortative mating (for an explicit comparison of different forms of assortative mating, see Felsenstein 1981). Both life cycles were used, and the minimum *s* values for speciation were found as described above.

To produce a version of the canonical two-trait (C2T) model, it is necessary to decouple mating from host choice, and have it depend instead on a second trait. This was done by keeping the assumptions of the above one-trait model, but letting mating take place in mating ''arenas,'' instead of on the hosts. A second set of *r* loci controlled choice of mating arena in similar fashion to the control of host choice, with each + allele increasing the probability of choosing mating arena I and decreasing the probability of choosing arena II. Mating arena preference thus acts as an assortative mating trait in this model, but one that is decoupled from the trait under disruptive selection. Only one life cycle was considered in this model, with selection after host choice: newborns → dispersal and host choice → selection on host → dispersal and mating arena choice → random mating within arenas → zygotes → meiosis → newborns. This life cycle is not meant to correspond to that of any known organism; its purpose is to produce a version of the C2T model for comparison with the results of the other models.

In the C2T model, speciation can occur only if linkage disequilibrium is initially present between the viability/host-preference loci and the mating arena preference loci. Therefore, to estimate minimum *s* values for the entire speciation process, a small amount of linkage disequilibrium was introduced by having 1% of the initial population consist of equal proportions of the two genotypes with all + or all − alleles, with the remaining 99% of the population as above (i.e. no linkage disequilibrium, all allele frequencies 0.5). Reducing the initial proportion of pure genotypes to 0.1% had only a trivial effect on minimum *s* values. Speciation completion was studied as in the other models.

The above assumptions were designed to make the three models as similar as possible, except for the relationship between assortative mating and disruptive selection. The models have the same fitness profiles insofar as is possible: in particular, genotypes with *i* host-preference/viability loci (1T and C2T model), or with *i* host-preference and *i* viability loci (Bush model), have the same fitness in all three models. In addition, the form of assortative mating is identical in all three models. Applicability of the phenotypic method to the 1T and C2T models is discussed in the Appendix.

Minimum *s* values for speciation are shown in Figure 4 for the 1T model with *n* loci, for the Bush model with equal numbers of host-preference and viability loci ($n = m$), and for the C2T model with equal numbers of host-preference/viability and mating arena preference loci ($n = r$). The latter two constraints make the three models equivalent except for the relationship between assortative mating and viability selection.

In terms of the minimum selection coefficients for speciation, the models rank C2T (circles in Fig. 4) ≫ Bush (tri-

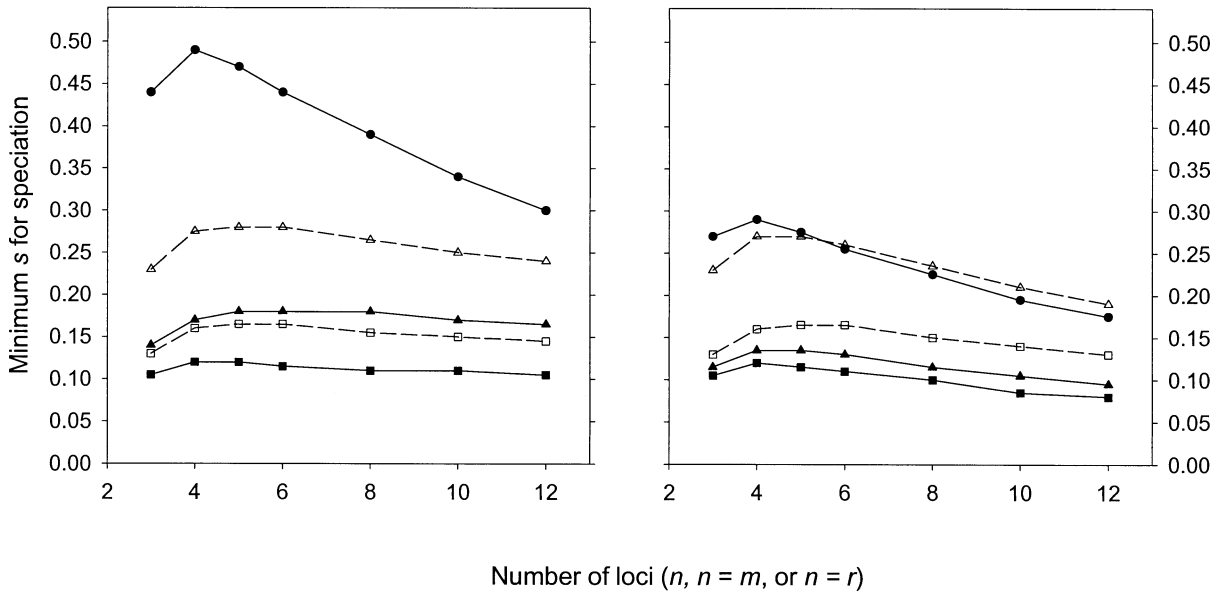

FIG. 4. Comparison of results of the three speciation models. On the x-axis is the number of loci; this is n in the one-trait (1T) model, $n = m$ in the Bush model, and $n = r$ in the canonical two-trait (C2T) model. Circles, C2T model; triangles, Bush model; squares, 1T model. Solid lines, selection after host choice; dashed lines, selection before host choice. Left, minimum s for entire speciation process; right, minimum s for speciation completion only.

angles) > 1T (squares). With selection after host choice (solid lines), selection coefficients must be one-third to two-thirds higher in the Bush model than in the 1T model for speciation to occur (Fig. 4, left). The difference in minimum s values for speciation completion between the Bush and 1T models is comparatively slight (Fig. 4, right). With selection before host choice (dashed lines), the differences in minimum s values between the Bush and 1T models are greater, but still well under two-fold. In contrast, selection coefficients must be nearly two-fold to over three-fold higher in the C2T model than in the Bush model for both the entire speciation process (Fig. 4, left) and speciation completion only (Fig. 4, right).

In terms of sensitivity to initial conditions, the models also rank C2T > Bush > 1T. In the 1T model, there is usually little difference between the minimum s values for the entire speciation process (Fig. 4, left) and those for speciation completion only (Fig. 4, right). For example, with $n = 6$ and selection after host choice, the respective thresholds are $s = 0.115$ and $s = 0.110$. For the Bush model, the corresponding thresholds are $s = 0.180$ and $s = 0.130$, whereas for the C2T model, they are $s = 0.44$ and $s = 0.255$.

DISCUSSION

The results of the analysis of the Bush model show that mating on the host coupled with selection for different genotypes on different hosts can lead to sympatric speciation under a fairly broad range of parameters. Speciation can occur when the fitness of intermediate, generalist genotypes is around one-half or more of that of the pure, host specialist genotypes. Contrary to the verbal model of Bush (1975), selection coefficients on individual viability loci do not need to be near one. The critical selection coefficients vary with the number of preference and viability loci, being lower with relatively few of the former and many of the latter. However, for a given amount of fitness difference between specialist and generalist genotypes (or between the two specialists on the host of one of them), speciation is more likely when the difference depends on relatively few loci. In addition, speciation occurs more readily when selection acts after individuals have chosen a host than when it acts on individuals whose host has been chosen for them by their parents. Possible reasons for this effect are discussed below.

Comparison to Previous Analyses of the Bush Model

Speciation in the version of the Bush model considered here is by no means automatic, because progress towards speciation is continually opposed by recombination between the multiple host-preference loci. Recombination regenerates individuals with $+-$ combinations of preference alleles, and hence with ambivalent host preference. In contrast, because similar recombination cannot happen in models with only a single preference locus (Diehl and Bush 1989; Kawecki 1996, 1997), such models reveal no minimum strength of selection for progress toward speciation. The progress toward speciation that can occur in these models, however, is much more limited than in the models considered here. In the haploid model of Diehl and Bush (1989), the preference locus had two alleles, which conferred equal and opposite host preference (e.g., A and a individuals choose host 1 with probabilities 0.7 and 0.3, respectively). Because of the frequency-dependent selection on habitat preference that occurs under soft selection, the two preference alleles quickly reached frequencies of 0.5, regardless of the strength of selection on the viability loci. After this point, no further evolution of host preference (and therefore of assortative mating) could occur. For this reason, it is questionable whether Diehl and Bush's model should be considered a model of speciation at all. In fairness, however, their main goal was to examine buildup

of linkage disequilibrium between the host-preference locus and the fitness loci; they showed that such linkage disequilibrium automatically develops. In contrast, in the influential two-trait model of Felsenstein (1981), in which the AM locus does not affect host preference, chance linkage disequilibrium between the AM locus and the viability loci decays unless the selection on the latter is very strong. Extrapolating from their results, Diehl and Bush correctly argued that the Bush model results in speciation more readily than the canonical two-trait model.

Kawecki (1996, 1997) considered a host-preference locus with three alleles, one conferring no preference, and each of the others conferring preference for one of the hosts. The assumption of three alleles means that the dynamics of preference are not entirely determined by frequency-dependent selection. Kawecki showed that if, at a second locus, beneficial (1996) or deleterious (1997) mutations with host-specific effects occur, the preference-conferring alleles will invade a population fixed for the neutral preference allele, no matter how weak the selection at the second locus. This can be viewed as progress towards speciation, in that increases in frequency of the preference-conferring alleles are accompanied by reduction of the level of gene flow among hosts. (An analogous result was obtained here in analysis of the exploratory model; as long as $s > 0$ at the viability loci, variation in host preference was favored). Completion of speciation, however, would likely require substitutions of preference alleles at multiple loci, followed by buildup of complete linkage disequilibrium among the multiple preference loci. Although there is no minimum strength of selection for the former process, the results presented here show that there is for the latter.

Kawecki (1996, 1997) also showed that an allele reducing migration between hosts is favored in his models (cf. Maynard Smith 1966; Balkau and Feldman 1973). This is an example of reproductive isolation evolving by a ''one-allele'' mechanism (Felsenstein 1981), in which the same allele is substituted in the incipient species. In contrast, all of the models considered in this paper are ''two-allele'' models, because the incipient species become fixed for different alleles at the AM loci. Although one-allele models work well in theory, genetic variation of the two-allele type appears to be more prevalent in populations (Felsenstein 1981; Berlocher and Feder 2002; Kirkpatrick and Ravigné 2002). In addition, one-allele models have the odd property of inevitably leading to sympatric speciation (Felsenstein 1981).

The comparison of alternative life cycles, with selection either before or after host choice, is another novel feature of the work reported here; Diehl and Bush (1989) and Kawecki (1996, 1997) considered only the life cycle with selection before host choice. Because in many insects selection occurs after as well as before host choice, it is important to consider both life cycles. In both the Bush and 1T models, the conditions for speciation were more stringent when selection acts before rather than after host choice. There appear to be two distinct causes of this difference (J.D. Fry, unpubl. data). First, in the life cycle with selection before host choice, fitness profiles are flatter for a given s than in the life cycle with selection after host choice. To take an extreme example, consider the fitness of the genotype with only + alleles. When selection occurs after host choice, these individuals will always choose host A, where their fitness will be 1. In contrast, with selection before host choice, some of these individuals will develop on host B, on which their fitness is low, because this host was chosen by their parents. Second, when selection occurs after host choice but before mating, the effective level of gene flow between genotypes that prefer different hosts is reduced compared to the life cycle with selection before host choice, in which mating immediately follows host choice.

Comparison of the Bush Model to the One-Trait and Canonical Two-Trait Models

One of the main goals of the present work was to compare the Bush model with the one-trait and canonical two-trait models, the most extensively analyzed models of sympatric speciation in the literature. These models differ in how disruptive selection is transmitted to the assortative mating trait (cf. Rice and Hostert 1993; Kirkpatrick and Ravigné 2002). In the one-trait model, genotypes with extreme and intermediate values of the AM trait always have the highest and lowest viability, respectively; disruptive selection in this case can be said to be direct. In the canonical two-trait model, the AM trait has no effect on viability, but can become genetically associated with the selected trait, resulting in indirect disruptive selection. The situation in the Bush model is more complex. In a population at linkage equilibrium with allele frequencies of one-half, host-preference genotypes all have the same marginal viability, averaged over genotypes at the viability loci. Conditional on viability genotype, however, host preference affects viability (either of an individual or its offspring, depending on the life cycle), with + preference alleles leading to higher (lower) viability in genetic backgrounds with mostly + (−) viability alleles, respectively. In other words, initially there is no disruptive selection on preference, but there is correlational or epistatic selection involving preference and viability loci. However, once linkage disequilibrium between the two types of loci builds up (as it will inevitably), host preference will affect mean viability, in such a way that genotypes with extreme preference will be favored. This can be appreciated by considering the effect of hypothetical allelic substitutions. Changing a + preference allele to a − allele in genotypes with all + preference alleles would reduce their average viability, because more than one-half of their viability alleles will be +, on average; the substitution will therefore increase the probability they choose the host to which they are poorly adapted. A similar substitution changing mating preference in the canonical two-trait model would have no effect on viability. Therefore the Bush model has an element of direct selection on the AM trait, unlike the C2T model, but resembles the C2T model in that linkage disequilibrium must be present for the AM trait to experience net disruptive selection (direct or indirect).

It has been widely recognized that direct disruptive selection on an AM trait, as in the one-trait model, results in sympatric speciation more readily than selection on a second trait that has a chance association with the AM trait, as in the canonical two-trait model (Rice and Hostert 1993; Kirkpatrick and Ravigné 2002). The large difference in thresholds for speciation between the versions of the 1T and C2T models

analyzed here support this conclusion. In contrast, the Bush model has not previously been compared to the 1T and C2T models in a multilocus framework. The above comparison of the types of selection operating in the three models suggests that the speciation thresholds in the Bush model should be lower than those in the C2T model, but greater than those in the 1T model. My results confirm this prediction, but show that thresholds for speciation in the Bush model are closer to those in the 1T model than those in the C2T model. This may stem from the quasi-direct nature of disruptive selection on host-preference loci in the Bush model, as discussed above.

Although speciation occurs more easily in the 1T model than in the other two models, fairly strong selection ($s > 0.10$; Fig. 4, left) is still required. Apparently, recombination is a significant obstacle to speciation in the 1T model, as well as in the other models. Indeed, the following argument shows that reducing recombination in the 1T model will make the conditions for speciation less stringent. Consider the 1T model with an even number of loci ($n = 2x$), each with selection coefficient s. Now suppose that the model is altered slightly to reduce recombination; the new model has x pairs of linked loci, with zero recombination between pair members, and free recombination among pairs. This model is in effect a model with x triallelic loci, with alleles $++$, $--$, and $+-$ or $-+$. It can easily be shown that the latter allele will be eliminated, leaving the other two to segregate. When this happens, the model becomes like the original biallelic 1T model, except with half as many loci and greater selection coefficients, s', on each one, where $s' = 2s(1 - s)$. From Fig. 4, it is apparent that there is always a range of s values where speciation does not occur in the model with $2x$ loci, but does occur in the reduced-recombination version with x loci (e.g., $n = 12$ and $s = 0.08$ in the original model becomes $n = 6$ and $s' = 0.15$ in the reduced-recombination model). Thus, although some authors have implied that recombination creates an obstacle to sympatric speciation only in two-trait models (Via 2001; Berlocher and Feder 2002), it is an obstacle in one-trait models as well. (Similarly, in the Bush model, it is not only recombination between preference loci and viability loci that impedes speciation, but recombination between loci of the same type. This is the apparent reason that increasing the number of preference or viability loci—while holding the total viability difference constant—was found to make speciation less likely). Because of the central role of recombination in opposing speciation, mechanisms that reduce recombination, such as chromosomal inversions, should increase the likelihood of speciation (Ortíz-Barrientos et al. 2002).

The three models considered here also differ strongly in their sensitivity to initial conditions. In the 1T model, it is almost as easy for sympatric speciation to occur in the first place as it is for the two species to remain separate when speciation is nearly complete. In the Bush model, the conditions for the entire speciation process are somewhat more stringent than those for speciation completion only, whereas in the C2T model, there is a large difference between the two sets of conditions (recall that if linkage disequilibrium is absent, speciation cannot occur at all in this model). These results make intuitive sense. In the C2T model, the strength of selection preserving linkage disequilibrium (LD) between the selected trait and the AM trait is an increasing function of the LD itself, going to zero as the LD goes to zero. The model is therefore more effective as a model of reinforcement, to which it has been compared (Howard 1993), than as a model of sympatric speciation. In contrast, in the 1T model, the combination of direct disruptive selection and assortative mating is nearly as effective in building up linkage disequilibrium as in preserving it.

Implications of Low Sensitivity to Initial Divergence of the One-Trait and Bush Models

The small difference between conditions for speciation initiation and speciation completion in the 1T model (and for some parameter combinations in the Bush model) has an important implication. Consider a pair of sympatric species that are reproductively isolated only by host or habitat preference. One of the arguments against putative cases of sympatric speciation is that speciation probably occurred in allopatry, with the two species coming into secondary contact (e.g., Mayr 1963; Futuyma and Mayer 1980). However, if *current* reproductive isolation depends only on host or habitat isolation, then this argument would lose some of its force: if selection is strong enough to keep the species separate, then there is a good chance that they could have speciated sympatrically.

The above argument may apply to the sympatric ''races'' of the pea aphid *Acyrthosiphon pisum*, one of which is affiliated with clover, and the other with alfalfa (Via 1991). Host fidelity provides an important source of premating isolation between the races (Via 1999; Caillaud and Via 2000), which readily mate in the laboratory to produce fertile hybrids (Via et al. 2000), suggesting that other forms of reproductive isolation may be weak. Furthermore, a QTL mapping study (Hawthorne and Via 2001) gave evidence for pleiotropy (or tight linkage) of host preference and host-specific fecundity. Therefore, the situation in the aphids may be better approximated by the 1T model than the Bush model, in which preference and fitness are controlled by unlinked loci. The cyclical parthenogenesis of the aphids should also increase the likelihood of sympatric speciation, by reducing the amount of recombination relative to selection. These considerations suggest that the genetic requirements for sympatric speciation by host isolation may be met in the aphids, although they give no evidence on whether the original divergence occurred in sympatry or allopatry.

The variation in sensitivity to initial conditions of the three models provides evidence against a claim made by the authors of a recent review of models of speciation by natural or sexual selection (Kirkpatrick and Ravigné 2002). Kirkpatrick and Ravigné classified published models by five criteria, including the type of selection on the AM trait (direct vs. indirect), initial degree of divergence (high or low), degree of geographic overlap (sympatric vs. parapatric vs. allopatric), and the genetic basis for reproductive isolation (one-allele vs. two-allele). The authors pointed out that the five criteria result in many possible combinations, less than half of which have been modeled, but went on to suggest that future analyses of particular combinations would contribute little: ''Modeling

every recombinational possibility of ingredients is an efficient algorithm for generating publications (and pulping trees) but not a good way to discover general principles'' (Kirkpatrick and Ravigné 2002, p. S28). This conclusion was apparently based on the authors' belief that ''[The] five elements appear to operate largely independently of one another'' (p. S22). If this claim were correct, then Kirkpatrick and Ravigné's point would be well taken, because the behavior of a model with a particular set of combinations could be accurately predicted without having to actually analyze the model. The results presented here, however, indicate that type of selection and the initial degree of divergence do not act independently, even approximately. Furthermore, it is not clear how Kirkpatrick and Ravigné arrived at the conclusion that the various elements act independently. To determine whether a particular pair of elements interacts, it is necessary to model all four (or more) combinations of the pair, holding other model details constant. Kirkpatrick and Ravigné's table 1 reveal that this has seldom been done.

Nonetheless, Kirkpatrick and Ravigné are correct in noting that the literature on speciation has become unnecessarily fragmented, due to the tendency of theorists to consider only one or two of the possible combinations of the elements at a time. The results reported here show that some unexpected insights can emerge from exploring how two or more of the elements interact.

Acknowledgments

I thank J. Feder, J. Jaenike, T. Kawecki, A. Kondrashov, M. Noor, H. A. Orr, and two anonymous reviewers for helpful comments. Supported by National Science Foundation grant DEB-0108730.

Literature Cited

Balkau, B. J., and M. W. Feldman. 1973. Selection for migration modification. Genetics 74:171–174.

Barton, N. H., and M. Shpak. 2000. The stability of symmetric solutions to polygenic models. Theor. Popul. Biol. 57:249–263.

Berlocher, S. H., and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu. Rev. Entomol. 47:773–815.

Bulmer, M. G. 1985. The mathematical theory of quantitative genetics. Clarendon, Oxford, U.K.

Bush, G. L. 1975. Sympatric speciation in phytophagous parasitic insects. Pp. 187–206 *in* P. W. Price, ed. Evolutionary strategies of parasitic insects and mites. Plenum, New York.

Caillaud, M. C., and S. Via. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. Am. Nat. 156:606–621.

Carroll, S. P., and C. Boyd. 1992. Host race radiation in the soapberry bug: natural history with the history. Evolution 46: 1052–1069.

Carroll, S. P., H. Dingle, and S. P. Klassen. 1997. Genetic differentiation of fitness-associated traits among rapidly evolving populations of the soapberry bug. Evolution 51:1182–1188.

Clarke, C. A., C. G. C. Dickson, and P. M. Sheppard. 1963. Larval color pattern in *Papilio demodocus*. Evolution 17:130–137.

Darwin, C. 1859. The origin of species by means of natural selection. Murray, London.

Dickinson, H., and J. Antonovics. 1973. Theoretical considerations of sympatric divergence. Am. Nat. 107:256–274.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Diehl, S. R., and G. L. Bush. 1989. The role of habitat preference in adaptation and speciation. Pp. 345–365 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, MA.

Doebeli, M. 1996. A quantitative genetic competition model for sympatric speciation. J. Evol. Biol. 9:893–909.

Doebeli, M., and U. Dieckmann. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am. Nat. 156:S77–S101.

Endler, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton, NJ.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Nature 336:61–64.

Feder, J. L., S. B. Opp, B. Wlazlo, K. Reynolds, and W. Go. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. Proc. Natl. Acad. Sci. USA 91: 7990–7994.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35:124–138.

Filchak, K. E., J. B. Roethele, and J. L. Feder. 2000. Natural selection and sympatric divergence in the apple maggot. Nature 407:739–742.

Fry, J. D. 1992. The mixed-model analysis of variance applied to quantitative genetics: biological meaning of the parameters. Evolution 46:540–550.

Futuyma, D. J., and G. C. Mayer. 1980. Non-allopatric speciation in animals. Syst. Zool. 29:254–271.

Harrison, R. G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. Pp. 19–31 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, New York.

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904–907.

Howard, D. J. 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. Pp. 46–69 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, New York.

Jaenike, J., and R. D. Holt. 1991. Genetic variation for habitat preference: evidence and implications. Am. Nat. 137:S67–S90.

Johnson, P. A., F. C. Hoppensteadt, and G. L. Bush. 1996. Conditions for sympatric speciation: a diploid model incorporating habitat fidelity and non-habitat assortative mating. Evol. Ecol. 10:187–205.

Jones, C. D. 1998. The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. Genetics 149:1899–1908.

Karowe, D. N. 1990. Predicting host range evolution: Colonization of *Coronilla varia* by *Colias philodice* (Lepidoptera: Pieridae). Evolution 44:1637–1647.

Kawecki, T. J. 1996. Sympatric speciation driven by beneficial mutations. Proc. R. Soc. Lond. B 263:1515–1520.

———. 1997. Sympatric speciation driven by deleterious mutations with habitat-specific effects. Evolution 51:1751–1763.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159: S22–S35.

Kondrashov, A. S. 1983a. Multilocus model of sympatric speciation. I. One character. Theor. Popul. Biol. 24:121–135.

———. 1983b. Multilocus model of sympatric speciation. II. Two characters. Theor. Popul Biol. 24:136–144.

———. 1986. Multilocus model of sympatric speciation. III. Computer simulations. Theor. Popul. Biol. 29:1–15.

Kondrashov, A. S., and F. A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400:351–354.

Kondrashov, A. S., L. Y. Yampolsky, and S. A. Shabalina. 1998. On the sympatric origin of species by means of natural selection. Pp. 90–98 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, New York.

Mallet, J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294–299.
Maynard Smith, J. 1966. Sympatric speciation. Am. Nat. 100: 637–650.
Mayr, E. 1963. Animal species and evolution. Belknap Press, Cambridge, MA.
Messina, F. J. 1991. Life-history variation in a seed beetle—adult egg-laying vs. larval competitive ability. Oecologia 85:447–455.
Moran, N. A. 1986. Morphological adaptation to host plants in *Uroleucon* (Homoptera: Aphididae). Evolution 40:1044–1050.
Ortíz-Barrientos, D., J. Reiland, J. Hey, and M. A. F. Noor. 2002. Recombination and the divergence of hybridizing species. Genetica 116:167–178.
Rausher, M. D., and R. Englander. 1987. The evolution of habitat preference II. Evolutionary genetic stability under soft selection. Theor. Popul. Biol. 31:116–139.
Rice, W. R. 1984. Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. Evolution 38:1251–1260.
Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47: 1637–1653.
Scharloo, W. 1970. Reproductive isolation by disruptive selection: did it occur? Am. Nat. 105:83–88.
Schliewen, U. K., D. Tautz, and S. Pääbo. 1994. Sympatric speciation suggested by monophyly of Crater Lake cichlids. Nature 368:629–632.
Shpak, M., and A. S. Kondrashov. 1999. Applicability of the hypergeometric phenotypic model to haploid and diploid populations. Evolution 53:600–604.
Thoday, J. M., and J. B. Gibson. 1962. Isolation by disruptive selection. Nature 193:1164–1166.
———. 1970. The probability of isolation by disruptive selection. Am. Nat. 104:219–230.
Via, S. 1991. The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. Evolution 45:827–852.
———. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. Evolution 53:1446–1457.
———. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16:381–390.
Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. Evolution 54:1626–1637.
Wolfram, S. 1996. The Mathematica book, 3rd ed. Wolfram Media/ Cambridge Univ. Press, Champaign, IL.

Corresponding Editor: M. Noor

APPENDIX

Applicability of Phenotypic Method

The key requirement for the method of following phenotype frequencies to work is that within phenotypes, frequencies of all genotypes are equal. For simplicity, I will term this the state of ''equiprobability.'' The equiprobable state implies that (1) within phenotypes (and therefore in the population as a whole), frequencies of + alleles are the same at each locus of a given type; and (2) within phenotypes, there is no linkage disequilibrium. If recombination is free, allelic effects are equal, and selection and assortative mating depend only on phenotypes, then a population in the equiprobable state will remain in that state, in the absence of random perturbations of genotype frequencies due to drift or mutation (Shpak and Kondrashov 1999). If a random perturbation of genotype frequencies occurs, the population may either evolve back to the equiprobable state, or away from it. The latter occurs in models of stabilizing selection, which evolve toward fixation of + alleles at some loci and − alleles at others (Bulmer 1985; Barton and Shpak 2000). In contrast, in models of sympatric speciation, there is a combination of disruptive and negative frequency-dependent selection, which together are favorable for maintaining the equiprobable state (Shpak and Kondrashov 1999; Barton and Shpak 2000).

TABLE A1. Results of deterministic iterations of 64 genotype frequencies for six loci. Random starting frequencies were generated as described in the text, with moderate initial linkage disequilibrium ($x = 0.8$). A total of 25 replicates were run for each parameter combination.

Model	n	m	s	No. of generations[1]	Mean (max) variance[2] $\times 10^{-4}$	Speciation?[3]
Bush	4	2	0.20	205	0.32 (0.56)	0/25
	4	2	0.40	31	11.7 (15.3)	25/25
	3	3	0.05	1157	0.49 (0.80)	0/25
	3	3	0.10	252	0.52 (0.72)	0/25
	3	3	0.15	88	1.24 (1.71)	25/25
One-trait	6	—	0.04	318	0.08 (0.10)	0/25
	6	—	0.08	111	0.12 (0.18)	0/25
	6	—	0.12	48	0.31 (0.46)	25/25

[1] Mean number of generations until all allele frequencies reached 0.5 ± 0.01.
[2] Mean and maximum of the variance of genotype frequencies within phenotypes, at the point when all allele frequencies reached 0.5 ± 0.01. Initial means were 460×10^{-4} for $n = 4$ and $m = 2$; 400×10^{-4} for $n = 3$ and $m = 3$; and 130×10^{-4} for $n = 6$.
[3] Proportion of replicates in which speciation occurred.

As described here, by numerical iterations of the complete set of 64 genotype frequencies for six unlinked loci, I confirmed that the equiprobable state is achieved in the Bush and one-trait (1T) models (with $n + m = 6$ and $n = 6$, respectively) from a wide range of initial genotype frequencies. For the canonical two-trait (C2T) model, the equiprobable state is not always achieved. Nonetheless, the phenotypic method should give an accurate description of evolutionary outcomes under this model if one additional assumption is made (see below).

To determine whether the equiprobable state would be reached given initial genotype frequencies departing strongly from equiprobability, vectors of initial frequencies, $\mathbf{p_0}$, were generated by taking the weighted sum of two random length 64 vectors: $\mathbf{p_0} = x\,\mathbf{v_1} + (1 - x)\,\mathbf{v_2}$. The vector $\mathbf{v_1}$ produces random departures from equal allele frequencies, whereas $\mathbf{v_2}$ produces random departures from linkage equilibrium. To generate $\mathbf{v_1}$, the six allele frequencies were chosen at random from a uniform distribution on 0,1; genotype frequencies were then generated under linkage equilibrium. To generate $\mathbf{v_2}$, 63 numbers were drawn at random from a uniform distribution on 0,1, and ranked from smallest to largest: $x_1, x_2, \ldots, x_{63}$. The elements of $\mathbf{v_2}$ were then calculated as $x_1, x_2 - x_1, \ldots, x_{63} - x_{62}, 1 - x_{63}$. Starting with $\mathbf{p_0}$, genotype frequencies were iterated until all allele frequencies were between 0.49 and 0.51, at which time the average variance of genotype frequencies within phenotypes was calculated. This variance serves as a measure of whether the equiprobable state had been achieved; variances of about 0.001 or less (SD $\leq$ 0.03) imply approximate equality of frequencies. Variances were calculated after first normalizing frequencies within each phenotype to sum to one. Iterations were continued to determine the eventual outcome (speciation or not).

Bush and one-trait models

Results for the Bush and one-trait models (with selection after host choice) are shown in Table A1, for various selection coefficients and moderate initial linkage disequilibrium ($x = 0.8$). In all cases, allele frequencies converged to 0.5. When allele frequencies were close to 0.5, the variance of genotype frequencies within phenotypes was always low (less than 2×10^{-3}).

The six-locus program was also used to check the speciation thresholds found by the phenotypic method program, for low initial linkage disequilibrium ($x = 0.99$). A total of 25 sets of random starting frequencies were used for each selection coefficient. Speciation always occurred just above the threshold found by the phenotypic method program (e.g., $s = 0.14$ for $n = m = 3$ in the Bush model with selection after host choice; see Fig. 3), and never occurred just below the threshold (e.g., $s = 0.135$).

Canonical two-trait model

In this model, the combined preference/viability loci are expected to reach allele frequencies of 0.5, as in the 1T model. In contrast, the mating arena preference loci are not under correlational selection analogous to that on the host-preference loci in the Bush model, and therefore should not be expected to always equilibrate at 0.5. This was confirmed in 25 runs with three loci of each type, $s = 0.5$, and random starting frequencies with $x = 0.8$. In eight cases, speciation occurred, so that all allele frequencies reached 0.5. In the remaining 17 cases, however, speciation did not occur. In these cases, final allele frequencies of the preference/viability loci were 0.5, but those of the mating loci ranged from 0.12 to 0.87, about the same range as the initial values.

At first, these results would indicate that the phenotypic method cannot be used to study the dynamics of the C2T model. However, the method can be ''rescued'' by adding one additional assumption to the model: that the mating loci are under negative frequency-dependent selection, with equilibrium allele frequencies of 0.5. Once allele frequencies reach 0.5, this selection ceases to operate; therefore frequency-dependent selection does not need to be explicitly incorporated into the phenotypic method program, and the magnitude of the selection is irrelevant. Because the sole purpose of the C2T model is for comparison with the other models, the biological source of the frequency-dependent selection need not be considered. Unsurprisingly, adding frequency-dependent selection to the six-locus program resulted in the equiprobable state being achieved regardless of the starting frequencies (J. D. Fry, unpubl. data).