EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 58 **June 2004** **No. 6**

Evolution, 58(6), 2004, pp. 1145–1156

ACCUMULATING DOBZHANSKY-MULLER INCOMPATIBILITIES: RECONCILING THEORY AND DATA

JOHN J. WELCH

Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton BN1 9QG, Sussex, United Kingdom
E-mail: johnwe0@central.susx.ac.uk

Abstract.—Theoretical models of the accumulation of Dobzhansky-Muller incompatibilities (DMIs) are studied, and in particular, the framework introduced by Orr (1995) and a verbal model introduced by Kondrashov et al. (2002). These models embody very different assumptions about the relationship between the substitution process underlying evolutionary divergence and the formation of incompatibilities. These differences have implications for our ability to make inferences about the divergence from patterns in the relevant data. With this in mind, the models are investigated for their ability to account for three patterns evident in this data: (1) the asymmetrical nature of incompatibilities under reciprocal introgression; (2) the finding that multiple concurrent introgressions may be necessary for an incompatibility to form; and (3) the finding that the probability of obtaining an incompatibility by introgressing a single amino acid remains roughly constant over a wide range of genetic distances. None of the models available in the literature can account for all of the empirical patterns. However, modified versions of the models can do so. Ways of discriminating between the different models are then discussed.

Key words.—Dobzhansky-Muller incompatibilities, epistasis, reproductive isolation, substitution process.

Received September 2, 2003. Accepted February 9, 2004.

When different populations of the same species undergo evolutionary divergence, each accumulates a distinct set of alleles. In some cases, these alleles may be complementary, that is, having an effect on fitness that is benign in their own genetic background but deleterious in the alternative background. When such alleles are present, hybridization between the two populations may bring together alleles that are intrinsically incompatible, resulting in offspring with reduced viability or fertility. If multiple incompatibilities accumulate, then complete postzygotic reproductive isolation may be the result. In recognition of the pioneering work of Dobzhanksy (1937) and Muller (1942), these incompatible combinations of alleles are commonly known as ''Dobzhansky-Muller incompatibilities'' (DMIs), and understanding their formation has become an important component in our understanding of the origin of species (Orr 1995, 1996; Gavrilets 1997; Barraclough et al. 1998; Coyne and Orr 1998; Barton 2001; Turelli et al. 2001; Edmands 2002; Gavrilets 2003).

As data on the accumulation of DMIs has itself accumulated (see below), several authors have suggested that it might yield insights into the process of divergence (e.g., Palopoli and Wu 1994; Wade and Goodnight 1998; Johnson 2000; Orr 2001; Turelli et al. 2001; Kondrashov et al. 2002). If certain patterns in the DMI data were shown to be reliable signatures of particular modes of divergence (such as natural selection or random genetic drift), then this would offer the tantalizing prospect of novel tests able to resolve long-standing controversies in the literature. These include debates about the relative importance of drift and natural selection in speciation events or the manner in which adaptive evolution commonly proceeds (e.g., Coyne and Orr 1998; Coyne et al. 2000; Goodnight and Wade 2000; Turelli et al. 2001).

Of course, our ability to infer process from pattern in this way depends on whether theoretical predictions can be found that apply exclusively to particular modes of divergence. Contrary to this hope, a great deal of the theoretical progress within this field has yielded predictions that are, in important respects, entirely independent of the mode of divergence. This applies, in particular, to the simple and elegant theoretical framework introduced by Orr (1995). Orr's framework, described below, is notable in being agnostic about the evolutionary forces driving the substitutions that differentiate the populations (Dobzhansky 1937; Muller 1942; Orr 1995; Coyne and Orr 1998). Indeed this generality has allowed it to be extended in various ways (e.g., Orr and Orr 1996; Orr and Turelli 2001; Gavrilets 2003; Kondrashov 2003). Perhaps the greatest success of this kind of work has been in illuminating the causes of Haldane's rule, a long-established pattern in studies of speciation (for a review see Orr 1997).

Recently, however, Kondrashov et al. (2002) introduced a

new method for investigating the accumulation of DMIs. Their empirical findings were not in accord with the basic predictions of Orr's (1995) framework. To explain their results, these authors introduced an alternative verbal model. This model contrasts with Orr's framework insofar as it relies on strong assumptions about the substitution process. As such, Kondrashov et al. suggest that their empirical findings constitute strong evidence for a particular mode of divergence (explained in full below).

In the light of these considerations, it is important to ask whether the assumptions about divergence embodied in the model of Kondrashov et al. (2002) really are necessary to account for their empirical results. The present study attempts to answer this question, but within a broader context. The study begins with a brief, partial survey of the currently available data on DMI accumulation, of which Kondrashov et al.'s (2002) results form only a part. I then ask whether any of the available models can account for all of the patterns evident in this data. It is found that neither the model of Orr (1995) nor that of Kondrashov et al. (2002) can successfully account for all of these patterns. The same is found to apply to alternative models available in the literature. As a result of these findings, modifications to the models are investigated. Finally, our ability to make inferences about evolutionary divergence from patterns in DMI accumulation is examined in a critical light.

Overview of Empirical Work

The first kind of data relevant to the present study involves experimental or natural hybridization, and relates a measure of genetic divergence (typically electrophoretic distance) to some measure of postzygotic reproductive isolation (see Coyne and Orr 1998; Edmands 2002 and references therein). The largest such study, Coyne and Orr's comprehensive compilation of *Drosophila* data (1989, 1997, 1998), has been followed by similar studies with frogs (Sasa et al. 1998); Lepidoptera (Presgraves 2002); various bird taxa (Price and Bouvier 2002); and pigeons and doves (Lijtmaer et al. 2003). In each case, these studies showed a gradual increase in reproductive isolation with divergence. This strongly suggests that multiple DMIs may be necessary for complete isolation to occur. Furthermore, in each case the correlation between isolation and divergence was found to be very roughly linear. Despite this important finding, the data do not allow us to draw detailed conclusions about how the number of DMIs increases with divergence. This is chiefly because we have little idea how multiple DMIs might interact to influence overall fitness (Sasa et al. 1998; Orr and Turelli 2001; Edmands 2002). Only if we were sure that DMIs were interacting independently, on average, could we conclude that the number of DMIs was increasing linearly with divergence. Furthermore, limitations in the data available to some of the studies mean that the correlations must be viewed with caution. These limitations include phylogenetically nonindependent datapoints and measures of reproductive isolation that record only full sterility or inviability. In addition, the data are often very noisy (Coyne and Orr 1998; Edmands 2002; Price and Bouvier 2002).

While hybridization studies allow us to view broad patterns in the data, a complementary approach is to examine the genetic architecture of reproductive isolation in particular cases. Relevant studies involve the experimental introgression of portions of chromosome from one background to another and then an assessment of fitness in the resulting hybrid. Such studies have typically been carried out using pairs of *Drosophila* species, with a refinement of conventional backcross procedures (e.g., Dobzhansky 1937; Wu and Beckenbach 1983; Orr and Coyne 1989; Perez et al. 1993; for reviews, see Coyne and Orr 1998; Johnson 2000), although divergent strains of other model organisms have also been used (e.g., Elena and Lenski 2001; Ungerer et al. 2003). Such studies allow us to determine both the density of factors capable of causing DMIs and the way in which these interact.

Two clear patterns have emerged from this work. First, those studies that have involved reciprocal introgression have tended to show that DMI formation is asymmetrical. This means that, if a chromosome segment is found to cause inviability or sterility when introgressed from species 1 into species 2, then, in general, the reciprocal introgression, from species 2 into species 1, is likely to show no such effect. A particularly clear example of this is Wu and Beckenbach's (1983) study (extending work of Dobzhansky 1937) with the sibling species *D. pseudoobscura* and *D. persimilis*. But the pattern is now well established (e.g., Orr and Coyne 1989; Coyne and Orr 1998).

The second pattern to emerge from these introgression studies is even more striking. This is the finding, common to many studies, that DMI formation often seems to require the introgression of multiple chromosome segments. For example, Cabot et al. (1994) located two factors on the *D. sechellia* X chromosome that caused sterility when introgressed together into a *D. simulans* background. However, neither factor, when introgressed alone, caused any significant reduction in fitness. Such results, often involving more than two introgressed factors, are now widespread (e.g., Cabot et al. 1994; Davis et al. 1994; Palopoli and Wu 1994; Perez and Wu 1995; Orr and Irving 2001).

The third procedure for investigating the accumulation of DMIs combines, to a remarkable extent, the strengths of the previous two approaches, although it has unique limitations, too. Kondrashov et al. (2002; see also Schaner et al. 2001) used a method that relates records of genetically mediated disease in humans to nonhuman sequence data. Specifically, Kondrashov et al. (2002) compiled a database of nonsynonymous (amino-acid changing) mutations known to be pathogenic in humans. They then collected the sequence data for various nonhuman orthologs of the mutating genes. With this data, they found evidence that the very same mutations were carried as the wild type in certain nonhuman lineages. These were species, then, from which a DMI-causing introgression could, in theory, have been performed. Using a method that accounted for the certain absence of many pathogenic mutations from their dataset, Kondrashov et al. were able to estimate the fraction of the amino-acid divergence that would prove seriously deleterious if introgressed (one codon at a time) from the nonhuman into the human background. This fraction is equivalent to the probability of obtaining an DMI by introgressing a single nonsynonymous codon.

The findings of these authors were striking in two respects.

First, this probability was found to be rather high—typically around 0.1. Second, and most remarkably, the probability was found to remain roughly constant across a very large range of genetic and phylogenetic distances. Indeed, Kondrashov et al. (2002) used data that spanned about 5% to 50% amino-acid divergence and nonhuman sequences ranging from primates to invertebrates.

These, then, are some of the empirical findings that theoretical treatments of DMI accumulation should be able to accommodate. This study will concentrate, in particular, on three broad patterns: (1) the asymmetrical nature of DMI formation; (2) the finding that multiple, simultaneous introgressions may be required for a DMI to form; and (3) the finding that the probability of obtaining a DMI by introducing a single foreign amino acid remains roughly constant at large divergences.

The third pattern, the constant probability, is obviously the least well established, relying, as it does, on a single study. The reliability of this study depends, in turn, on the accuracy with which human pathogenic mutations have been identified (Kondrashov et al. 2002). The constant probability does agree well with our naive interpretation of the hybridization studies—that the linear increase in reproductive isolation reflects a linear increase in the number of DMIs—but, as explained above, we can have little confidence in this interpretation. As a result, in the absence of work replicating or contradicting the findings of Kondrashov et al. (2002), the present study will proceed as if this pattern is established (see Discussion).

When it comes to testing the theoretical predictions against these three empirical patterns, it will be helpful, repeatedly, to use a statistic that will be denoted P_m. To define this statistic, consider a pair of genetic backgrounds that have diverged from a common ancestor. From the genetic factors that distinguish the backgrounds, we choose m factors at random and introgress all m from one background to the other. P_m is defined as the probability that the scenario just described will lead to the formation of a DMI involving all m of the introgressed factors. Note that this definition explicitly excludes DMIs that might be formed by introgressing some subset of the m factors alone.

With this definition, we can express two of the empirical patterns in terms of P_m. First, when the introgressed factors are considered to be single nonsynonymous codons, Kondrashov et al.'s (2002) key result can be written as $P_1 \approx 0.1$ (regardless of the level of divergence). Second, the findings of Cabot et al. (1994) and others, that multiple introgressions may be required for a DMI to form, suggest that P_m must be nonzero when $m > 1$. Furthermore, the abundance of similar findings suggests that $P_2 \ll P_1$ should not hold; if $P_2 \ll P_1$ did hold, then it would be extremely uncommon to observe an incompatibility that required a double introgression when the two individual introgressions had no discernible effect. With these results in mind, Orr's (1995) framework is now introduced.

The Combinatorial Model

To understand many of the models discussed here, it is crucial to distinguish between those combinations of alleles that reached high frequencies and substantial copy number during the process of divergence and those that did not. Precisely because the former group reached high frequencies and copy numbers, it is safe to assume that they confer a reasonable level of fitness (Barton 2001). In contrast, the latter group may contain combinations of alleles that are intrinsically incompatible. We might say that the compatibility of the former group is proven, while that of the latter group is unproven. For brevity, in what follows, I will refer repeatedly to ''proven combinations'' (those that were present at high copy number and frequency during divergence) and ''unproven combinations'' (all of the others).

Bateson (1909), Dobzhansky (1937), and Muller (1942) were the first to show that unproven combinations of alleles can be formed by hybridization, even when the hybridizing populations have diverged through a process of single, sequential substitutions, each of which was unopposed by natural selection (Orr 1995, 1996; Gavrilets 1997; Johnson 2002). Dobzhansky (1937) proposed a two-locus model to demonstrate this idea, and Orr's (1995) framework is a multilocus extension of this model.

Consider a pair of coding sequences that have diverged from a common ancestor and now differ by D amino acids. Through hybridization, unproven combinations of amino acids may be formed. Under the assumptions of Orr's model, each unproven combination of n amino acids (where $n \leq D$), wherever present in the hybrid sequence, can potentially form a DMI. The total number of unproven combinations of size n that may be formed from the pair of sequences is denoted $C(D, n)$. Although each of these $C(D, n)$ unproven combinations may form a DMI, it is not certain to do so. Under the assumptions of the model, each potential incompatibility constitutes an actual incompatibility with a fixed probability. This probability is denoted p_n.

Under these assumptions, the expected number of DMIs involving n amino acids, $E[I_n]$, is simply:

$$E[I_n] = p_n C(D, n) \tag{1}$$

(Orr 1995). (The complete distribution of I_n is binomial, with the number of trials equal to $C[D, n]$, and the probability of success equal to p_n). To calculate $C(D, n)$, we must make two (relatively weak) assumptions about the process of divergence. First, we assume that the nonsynonymous divergence came about by single, sequential substitutions. Second, we assume that each of these substitutions took place at a unique site. Under these assumptions, we know that D substitutions must have taken place in total, and that $D + 1$ complete haplotypes will have been proven to confer high fitness. By similar considerations, the number of unproven combinations of n amino acids is found to be

$$C(D, n) = \binom{D}{n}(2^n - n - 1), \tag{2}$$

where we have used the binomial coefficient, $\binom{D}{n} = D!/[n!(D - n)!]$. (The appendix gives a brief derivation of eq. 2 and discusses some consequences of relaxing the assumptions.) Note that equation (2) differs from the corresponding expressions given by Orr (1995) and Gavrilets (2003) when $n > 2$. However, the differences do not affect the major conclusions of these authors. In addition, there is full agreement

when exclusively pairwise DMIs (with $n = 2$) are considered (Orr 1995; Orr and Turelli 2001).

The most important point about equations (1) and (2) is that, together, they predict that the number of DMIs will increase very rapidly with the divergence, D. Indeed, as pointed out by Orr (1995), the number of n-fold incompatibilities will increase roughly as D^n. Orr dubbed this striking pattern the ''snowball effect.''

A second important point to notice about the above treatment is that, following Orr (1995) and Orr and Turelli (2001), dominance has been ignored. This is quite appropriate for our purposes, because Kondrashov et al.'s study (2002) uses both recessive and dominant pathogenic mutations (citing, e.g., Polymeropoulos et al. 1997; Schaner et al. 2001), and treats them interchangeably. In addition, many of the introgression studies mentioned above have used hemizygous sex chromosomes (the *Drosophila* X), or haploids. (Note, in addition, that deleterious interactions between maternally and paternally inherited amino acids at the same site are ruled out by the assumption that no multiple substitutions occur.)

Implications for Divergence

The simplicity of the framework above stems partly from the fact, noted by Muller (1942), that the accumulation of DMIs proceeds regardless of the distribution of the substitutions among the two lineages. Equation (2) will remain unchanged, for example, if all D substitutions take place in one lineage while the other remains in their common ancestral state. The second reason for the simplicity, as has been mentioned, is the relative independence of the model from the substitution process. Equations (1) and (2) assume only that the DMIs have not appeared at high frequency during the divergence of the populations. Apart from this, the results apply regardless of the evolutionary forces that brought about the substitutions. Of course, various assumptions about the process of divergence can and have been added to the model (Orr 1995; Orr and Orr 1996; Orr and Turelli 2001; Kondrashov 2003). But none of these affect equations (1) or (2), and these can be related directly to the data.

A consequence of these considerations is that, if the above framework is consistent with the data, there is little prospect of using DMI data to make inferences about the process of divergence.

Relation to Data

How well, then, does the model account for the empirical patterns noted earlier? To answer this question, we must calculate the statistic P_m. For this purpose, we require a quantity that is similar but not identical to $C(D, n)$ (eq. 2) and whose definition is unavoidably convoluted. Specifically, we are interested in unproven combinations that may be formed by the introduction, from one background to the other, of a particular set of m amino acids, where the m have been chosen at random from the D amino acids that distinguish the backgrounds. Furthermore, given our definition of P_m, we must count only those combinations that include all m of the introgressed amino acids. With this in mind, $C(D, n, m)$ is defined as the expected number of unproven combinations of n amino acids that may be formed by introgressing a particular randomly chosen set of m divergent amino acids, counting only combinations that include all m of the introgressions. (This definition implies that $n \geq m$.) Under the assumptions listed above, it is found that

$$C(D, n, m) = \frac{\binom{D}{n}}{\binom{D}{m}}\left[\binom{n}{m} - 1\right] \tag{3}$$

(see the appendix for a derivation). The expected number of DMIs formed in this way is then $E[I_{n,m}] = p_n C(D, n, m)$ We can now find P_m, the probability that a DMI of any size (any value of n) will be formed. If $E[I_{n,m}] \ll 1$, this probability is approximately

$$P_m \simeq 1 - \prod_{n=m}^{D} (1 - E[I_{n,m}]). \tag{4}$$

It follows from equation (4) that Orr's framework is not consistent with Kondrashov et al.'s (2002) finding that P_1 remains roughly constant over a range of divergences. Instead, equations (3) and (4) together predict that P_1 should increase at least linearly with genetic distance—a manifestation of the snowball effect. To see this clearly, consider the situation where only pairwise incompatibilities, involving $n = 2$ amino acids, can form. In this case, $p_n = 0$ for $n \neq 2$, and we find

$$P_1 \simeq E[I_{2,1}] = \frac{p_2}{2}(D - 1). \tag{5}$$

This factor alone shows a linear increase with D, and including higher-order incompatibilities (the $n > 2$ terms in eq. 4) can only make the increase with D more rapid.

In contrast, Orr's framework *can* account for the finding that DMIs tend be asymmetrical. Muller (1942) noted that under Dobzhansky's two-locus model (which is equivalent to setting $D = n = 2$ above) DMIs are guaranteed to be asymmetrical. However, as D increases, the outcomes of a pair of reciprocal introgressions become increasingly independent. As such, an asymmetrical DMI will be obtained with a probability roughly equal to $P_m(1 - P_m)$, while a symmetrical DMI will be obtained with probability P_m^2 (see also Orr 1995). If P_m is small, then asymmetrical DMIs should be more common. This prediction will prove to be common to all of the models examined here that are related to work of Bateson, Dobzhansky, and Muller.

The final empirical pattern is the finding that multiple concurrent introgressions may be necessary for a DMI to form. Here, the relation of the model to the data is more ambiguous. Orr's framework certainly predicts that such DMIs may exist. This follows immediately from the fact that P_m may be nonzero when $m > 1$ (eq. 4). This prediction is, in fact, built in to the model via the assumption that any unproven combination of n factors may form a DMI. Because n may be greater than two, the model allows for complex epistasis, that is, fitness interactions between more than two factors (Cabot et al. 1994). Furthermore, as pointed out by Orr (1995), the model predicts that such epistasis will be common. This is because the number of unproven combinations, $C(D, n, m)$,

will increase with n while $n < D/2$ (and because we do not expect individual DMIs to involve very large numbers of interacting elements, $n \ll D$ seems biologically reasonable). The same conclusion follows from arguments relating to ''possible evolutionary paths'' (Cabot et al. 1994; Orr 1995; see the appendix).

However, a DMI containing any number of interacting factors can be formed by introgressing just a single factor, and it does not follow that the introgression of multiple factors is a likely prerequisite for DMI formation. In other words, although P_m may increase with n (the total number of factors involved in the DMI), it does not follow that it will increase with m (the total number of introgressed factors involved). Indeed, equation (4) predicts just the opposite. To see this, consider again a single value of n (this time arbitrary). In this case, the ratio of P_1 and P_2 is found to be

$$\frac{P_1}{P_2} = \frac{(D-1)(n-1)}{(n+1)(n-2)}, \tag{6}$$

if $p_i = 0$ for $i \neq n$. So for $n \ll D$ (which, it has been suggested, is biologically reasonable), the model predicts that DMIs involving a single introgressed amino acid are far more likely to be detected than DMIs that require the concurrent introgression of two amino acids. The same pattern holds for other values of m, and in general, when small, P_m is roughly proportional to D^{n-m}. As a result, given a species pair for which $D \approx 10^5$ (e.g., Palopoli and Wu 1994; Orr and Turelli 2001), we would expect to see many thousands of DMIs requiring a single introgression for every DMI that requires two or three introgressions. This seems difficult to reconcile with the frequent discovery of such DMIs in the introgression studies (Coyne and Orr 1998; Johnson 2000).

It may be argued that the above quantities are not relevant to explaining the empirical results, because each of the introgressed factors in experiments such as those of Cabot et al. (1994) will almost certainly have coded for multiple divergent amino acids. To explain the results of these authors, it is required only that the DMI detected involved at least one amino acid from each factor. With this in mind, consider the introgression of m factors, each of which codes for a divergent amino acids. Let $P_m^{(a)}$ denote the probability that such an introgression will lead to a DMI involving at least one amino acid from each of the m factors. (Clearly, with this definition, $P_m \equiv P_m^{(1)}$.) Again, we assume that the DMI involves n amino acids. The probability $P_m^{(a)}$, for general a, is difficult to find analytically except in a very unhelpful form. However, supplementary material (available online at http://dx.doi.org/10.1554/03-502.1.s1) provides a derivation of expressions for $P_1^{(a)}$ and $P_2^{(a)}$ and shows that, to a very rough approximation,

$$\frac{P_1^{(a)}}{P_2^{(a)}} \approx \frac{D}{a(n-1)}, \qquad D \gg a, n. \tag{7}$$

This agrees qualitatively with the result for single factor introgressions, equation (6).

The Epistatic Selection Model

As mentioned above, to explain their finding of a constant P_1, a finding not in accord with equation (5), Kondrashov et al. (2002, pp. 14881–14882) presented an alternative verbal model. Like Orr's model, the model of Kondrashov et al. is consistent with the basic insights of Bateson, Dobzhansky, and Muller—indeed it is very close to a verbal description given by Muller (1942, pp. 87–88). However, in contrast to Orr's approach, Kondrashov et al.'s model relies on additional assumptions about the substitution dynamics.

In particular, Kondrashov et al.'s model assumes that pairs of DMI-causing substitutions follow each other in rapid succession (rapidly, that is, on the timescale of molecular evolution), and that this is due to the action of positive, epistatic selection. The selection is epistatic in the sense that the presence or absence of the first substitution in each pair determines the direction of selection acting at the second site. Specifically, before the first, *precursor substitution* has taken place, the second, *epistatically selected substitution* is selected against, and is thus prevented from occurring. After the precursor substitution, in contrast, the epistatically selected substitution is subject to positive selection and so will be driven rapidly to fixation. Under the assumptions of the model, the epistatically selected substitution may, in turn, alter the direction of selection at a third site, resulting in a chain of positively selected substitutions. Note that this positive selection is a crucial aspect of the model—the behavior would alter qualitatively if the precursor substitution merely released selective constraint at the second site (Kondrashov et al. 2002). Note, too, that the epistatic nature of the selection does not entail that any epistatic genetic variance be expressed.

Given the assumptions above, it follows that an introgression placing an epistatically selected substitution in a background lacking its precursor substitution will form a DMI. As such, the formation of DMIs and the substitution process are intimately linked.

Finally, note that the sequential nature of the substitutions is not a crucial aspect of the model. All of the relevant statistical properties would remain unaltered if certain pairs occurred simultaneously, as a result of the formation of a double mutant (Kondrashov et al. 2002).

Implications for Divergence

If the assumptions behind this model hold, then (as noted by Kondrashov et al. 2002) there follow a number of important implications for various evolutionary problems. These include implications for the substitution process in molecular evolution, stemming both from the high frequency of positive selection and from its epistatic nature (e.g., Gillespie 1984, 1991; Ohta 1997; Smith and Eyre-Walker 2002). These, in turn, have implications for phylogenetic reconstruction and the placing of evolutionary events in time (e.g., Bromham and Penny 2003). On a quite different point, if the development of postzygotic reproductive isolation is coupled to adaptive substitution in this way, then this calls into question the relevance of exclusively drift-based theories (e.g., Rice and Hostert 1993; Orr and Orr 1996; Gavrilets 1997, 1999; Wade and Goodnight 1998). In addition, the epistasis posited by the model would have implications for the evolution of sexual reproduction (e.g., Maynard Smith 1978; Peters and Lively 2000; Kondrashov and Kondrashov 2001).

Clearly then, it is important to ask if this model can satisfactorily account for the known data.

Relation to Data

To show that the verbal model is consistent with a constant P_1, it must be formalized. Luckily, this is trivial. Let us assume that with some fixed probability, q, each new substitution causes the selection pressure that brings about the subsequent substitution. Although the first substitution that occurred in each lineage cannot have been the result of such a process, every subsequent substitution may have been so caused (since epistatically selected substitutions may in turn bring about epistatic selection at another site). As such, given a genetic distance of D, we know that a maximum of $D - 2$ epistatically selected substitutions have occurred. Since, by assumption, each epistatically selected substitution creates an incompatibility, the number of DMIs is a binomially distributed random variable with the expected value $q \times (D - 2)$. This yields a value of P_1 as follows

$$P_1 = \frac{q(D-2)}{D}, \qquad D > 1. \tag{8}$$

Clearly, as D becomes large, $P_1 \simeq q$ and so the proportion of incompatibilities reaches a constant value, independent of D. As such, with $q = 0.1$, this model is consistent with Kondrashov et al.'s (2002) empirical findings.

Secondly, it is clear that Kondrashov et al.'s model, like Orr's (1995), predicts that DMIs will tend to be asymmetrical. This follows directly from the basic asymmetricality between the precursor and epistatically selected substitutions (the former, but not the latter, may be benign in all backgrounds). Since a fraction $q \simeq P_1$ of precursor substitutions will themselves have been epistatically selected, a fraction $(1 - P_1)$ of DMIs are expected to be asymmetrical. This result is identical to that from Orr's model.

Finally, and in contrast to Orr's model, it is clear that, under the epistatic selection model, a DMI may only form between a pair of factors—one from each background. As such, $P_m = 0$ when $m > 1$, and so the model cannot account for the finding that DMI formation may require multiple concurrent introgressions.

Two kinds of modifications could be made to the epistatic selection model to allow it to account for the findings of the introgression studies. First, we could simply relax the assumption that DMIs form between pairs of sites. This, however, would have the consequence that P_1 would not remain constant. The second kind of modification is a little more involved and is discussed in the following section.

The Epistatic Selection Model with Genetic Redundancy

As has been noted by several authors, the finding that multiple introgressions may be necessary for a DMI to form might not be due to complex epistasis (the interaction of multiple factors), but might instead by due to some kind of genetic redundancy (Cabot et al. 1994; Palopoli and Wu 1994; Orr 1995; Johnson 2000). A function carried out redundantly by a pair of genes will continue unaffected if either alone is inactivated, but will fail, causing a probable loss of fitness, if both are inactivated simultaneously. This would lead to fertility or viability acting as a threshold trait (Palopoli and Wu 1994; Johnson 2000).

The existence of genetic redundancy is well established, but there is little consensus as to how it might evolve. Nowak et al. (1997) discuss the classification of different types of redundancy and model various scenarios for its evolutionary maintenance. Here, we simply wish to establish that some kind of redundancy might be consistent with the DMI data. As such, the simplest possible extension of the epistatic selection model will be considered. This extension should be very roughly consistent with all of the scenarios considered by Nowak et al. (1997).

With this in mind, assume that, after an epistatically selected substitution has taken place, each subsequent substitution occurring in the lineage may redundantly carry out the function of its precursor substitution. Such a redundancy-causing substitution is assumed to occur with a fixed probability, r. The result of such redundancy is that an epistatically selected substitution will only cause a DMI in a background lacking both its precursor substitution and all of the redundancy-causing substitutions.

Unfortunately, the model just described requires that we track the substitutions that occur in each lineage separately. Supplementary Material 2 (available online) contains a full derivation, but in the main text we assume that an equal number of substitutions have occurred in each lineage. With the foregoing assumptions, it is found that

$$P_1 = \frac{q}{2D}\left\{\mathrm{D} - 2 + \frac{2}{r}[1 - (1 - r)^{(D-2)/2}]\right\}. \tag{9}$$

In this expression, q, as before, denotes the probability that a substitution brings about the epistatic selection pressure causing a further substitution, D denotes the divergence, and r denotes the probability that a substitution creates redundancy.

The behavior of equation (9) is clearest at small and large divergences. While $rD \ll 1$, equation (9) is approximately equal to the result without redundancy, equation (8). When $rD \gg 1$, on the other hand, equation (9) approaches the constant value $P_1 \simeq q/2$. In this case, to account for Kondrashov et al.'s finding that $P_1 \simeq 0.1$, it must assumed that at least 20% of the substitutions separating the lineages were the result of positive epistatic selection (since $q \simeq 2P_1$). This high rate is twice that required by Kondrashov et al.'s original model (for which $P_1 \simeq q$) but still well within some empirically obtained estimates (Fay et al. 2002; Smith and Eyre-Walker 2002). When r is very small (as seems biologically plausible) convergence to the constant value of P_1 is extremely slow. However, after the first few substitutions, P_1 begins a slow decline that would be difficult to detect in a noisy dataset.

As well as producing an approximately constant P_1, the incorporation of genetic redundancy allows this framework to accommodate the introgression results. As shown in Supplementary Material 2 (available online) for large D the following approximation can be derived:

$$P_m \simeq \binom{D}{m}^{-1}\frac{q}{r}, \qquad m > 1. \tag{10}$$

This is clearly nonzero, and so DMIs requiring multiple introgressions will be possible. However, the decline of equation (10) with m will be even more rapid than was the case for Orr's model. So, for example, at large divergences, $P_1/P_2 \simeq D^2r/4$ under this model (a result which can be compared with eq. 6).

The particulars of the model just presented are certainly debatable, but it serves to establish the principle that epistatic selection, combined with genetic redundancy (of some sort), can account for all of the empirically observed patterns. However, it is still not clear that this kind of model is necessary to explain the patterns. With this in mind, some alternative models are now examined.

The Russian Roulette Model

Gavrilets and Gravner (1997; see also Gavrilets 1997, 1999, 2003) introduced a novel framework for investigating the evolution of reproductive isolation. Under the assumptions of their model, coding sequences are assigned to one of two categories, high fitness and low fitness (these need not be fixed values, as is assumed by a few treatments, but can encompass a range of values). The assignment of each sequence to one of the categories takes place independently and at random. In particular, each sequence is assigned a high level of fitness, with a probability f, and a low level of fitness otherwise; that is, with probability $(1 - f)$. This procedure, loosely resembling Russian roulette, means that similar genotypes are no more likely to have similar levels of fitness than are highly divergent genotypes.

Despite this, if f is sufficiently large, chains of high-fitness genotypes, each differing by a single mutation, may be present (Gavrilets 1997, 2003). In this case, populations may diverge via the fixation of single, sequential substitutions, without ever passing through a state of low fitness. This description makes it clear that Gavrilets's model is consistent with the scenario described by Bateson (1909), Dobzhansky (1937), and Muller (1942). Indeed, like Orr's combinatorial model, the Russian roulette model is a multilocus extension of Dobzhansky's (1937) two-locus model. In addition, and also like Orr's model, the Russian roulette model makes no assumptions about what drives the substitutions leading to the divergence. (Although, as with Orr's model, assumptions about divergence can be included; e.g., Gavrilets 1999.)

Relation to Data

Under this model, a DMI will be formed if an introgression between two divergent high-fitness sequences results in a novel, low-fitness sequence. The model's simplicity makes it quick to calculate

$$P_m = \left[1 - 1 \Big/ \binom{D}{m}\right](1 - f). \tag{11}$$

Here, the term in the square brackets is the probability that the introgression recreates an unproven combination, that is, a combination not present at high frequency during divergence (see eq. A1 in the appendix). For $D \gg m$, this term approaches unity, and so $P_m \simeq (1 - f)$ independent of D and m. As a result, this model can account for both the constancy of P_1, and, since all values of P_m are roughly equal, for the frequent detection of DMIs requiring multiple concurrent introgressions. This is in contrast to all previous models.

The fraction of introgressions that are expected to be asymmetrical is also quick to calculate. For single-factor introgressions (for which $m = 1$) , this quantity is

$$1 - \left(\frac{D - 2}{D - 1}\right)(1 - f). \tag{12}$$

Because equation (12) is equal to unity when $D = 2$, it makes explicit Muller's (1942) finding that all DMIs will be asymmetrical in the two-locus case. In addition, it shows that, as with all previous models, the fraction of asymmetrical DMIs approaches $(1 - P_m)$ as D becomes large.

The Russian roulette model, then, can account for all three of the empirically observed patterns more successfully than can either the combinatorial model or the epistatic selection model, even in its modified form. However, there are two major problems with accepting this framework as an explanation for the data. First, to account for the quantitative finding of Kondrashov et al. (2002) that $P_1 \simeq 0.1$, we require $f \simeq 0.9$ (eq. 11). In other words, we must believe that 90% of randomly generated sequences confer high fitness. This seems highly implausible. Second, because fitness is an all-or-nothing affair under the model, it cannot account for the finding that reproductive isolation may require the formation of multiple DMIs, a conclusion that follows from the hybridization studies discussed above.

Despite these limitations, the model suggests that the finding of a constant P_1 may not require an explanation positing a particular mode of divergence; instead, it may be result from the consideration of sections of coding sequences as integrated wholes. In the following section, a modification of Orr's combinatorial model is introduced that incorporates this suggestion.

The Combinatorial Model with Saturation

The above treatment of Orr's (1995) model assumed that the incompatibility-causing interactions took place between individual amino acids. However, we gain a different perspective if it is assumed that DMIs form between combinations of some higher functional unit—perhaps binding sites or whole proteins. The significance of this change stems from the fact that the disruption of a given functional unit might be achieved by replacing many different amino acids. As a result, under these assumptions, multiple amino acid replacements (undertaken one at a time) might bring about essentially the same incompatibility. This modified picture is in line with the studies of genetic pathology cited by Kondrashov et al. (e.g., Polymeropoulos et al. 1997; Schaner et al. 2001), and, incidently, seems closer to Orr's (1995) original treatment.

To model this change in a succinct manner, we introduce a variable, K, that will represent functional divergence. If we consider a pair of orthologous sequences that carry out multiple functions, K can be roughly characterized as the number of functions carried out by one, but not both, of the sequences. The functional divergence is related to the amino acid divergence, D, insofar as functional change must be due to amino acid change. (Although a particular amino acid change

need not cause significant functional change, ensuring that $K \leq D$.) The number of potential DMIs will now depend on the number of unproven functional combinations that can be created, that is, on $C(K, n)$ rather than $C(D, n)$ (see eq. 2).

A final key assumption is that the functional divergence, K, has a maximum value, which we will denote $\hat{K}$; once this value has been reached, D may continue to increase but K cannot. What does this mean in less abstract terms? The model relies on the assumption that introgression between a pair of orthologous sequences may allow us disrupt some of the functions carried out by those sequences. The inclusion of $\hat{K}$ is tantamount to assuming that, after a certain level of sequence divergence, enough raw material is present to allow us to disrupt almost every function of the sequences in this way.

To complete the model, we must now decide how K relates to D before the saturation point ($K = \hat{K}$) has been reached. Here, stochasticity play a role, and so each substitution (while certain to increase D by one) leads to an increase in K with a certain probability. The particular functional form we choose will prove relatively unimportant, and so we take a simple linear relationship, assuming that a given substitution increases K with a probability equal to $1 - (K/\hat{K})$.

We first obtain an approximation for the expected number of DMIs that may be formed between two backgrounds differing by D amino acids. Using equations (1) and (2), we find

$$E[I_n] = p_n \sum_{k=0}^{\hat{K}} \Pr\{K = k; D, \hat{K}\} C(k, n) \approx p_n C(\hat{K}, n)[1 - e^{-D/\hat{K}}]^n. \quad (13)$$

(Supplementary Material 3, available online, gives the exact result and derivation.) Equation (13) behaves quite differently at small and large divergences. In the initial stages of divergence, when $D \ll \hat{K}$, the term in the square brackets can be approximated by $[D/\hat{K}]^n$, and so the number of DMIs is proportional to D^n, creating a snowball effect, just as with Orr's original model (eq. 1). At the other extreme, when divergence increases far beyond the saturation point, $D \gg \hat{K}$, the term in square brackets approaches unity, and so the number of DMIs asymptotically approaches a constant value. Interestingly, this agrees with an assertion by Muller (1942, p. 101; Johnson 2002).

According to the assumptions above, each of these DMIs may be constructed by individually introgressing multiple amino acids. Indeed, a particular unproven combination may be formed by an average of D/K single introgressions. This complication makes P_1 difficult to obtain. However, once genetic divergence is very high, $D \gg \hat{K}$, the functional divergence is almost certain to have reached its maximum, $K = \hat{K}$, and so, using equation (3), we have

$$P_1 \simeq 1 - \prod_{n=2}^{\hat{K}} [1 - p_n C(\hat{K}, n, 1)], \qquad D \gg \hat{K}. \quad (14)$$

This is a constant value, and so, in principle, a model like the one above could account for all of the empirical findings.

How plausible, then, are the assumptions behind this model? To answer this question, it is useful to consider the behavior of the combinatorial model at very large divergences, without some kind of saturation taking place. In this case, as a result of the snowballing, P_1 approaches unity rather rapidly. (Obtaining an n-fold DMI becomes very probable once $D^{n-1}p_n > 1$, for example.) Intuitively, however, it seems implausible that all single-amino-acid introgressions would cause severe fitness loss, even between the most divergent genotypes.

Since this intuition applies only at very large divergences, it does not follow that the model above provides a plausible account of Kondrashov et al.'s (2002) empirical results. To explain their results with equation (14), we must assume that the saturation point, $D \gg \hat{K}$, can be reached after as little as ~5% amino acid divergence (the smallest in their dataset). This seems most implausible. To refine intuition, a crude estimate of $\hat{K}$ can be obtained by adapting the method of Orr and Turelli (2001). These authors, assuming that all DMIs were pairwise ($n = 2$) , used a variant of equation (1) to estimate p_2 for various hybridizations. Since, under the saturating model, equation (1) is expected to apply for a recently diverged species pair, estimates of p_2 obtained in this way can be used in equation (14) to estimate $\hat{K}$. Relevant data can be obtained from Orr and Irving's (2001) study of the hybridization between the recently diverged Bogota and U.S. subspecies of *Drosophila pseudoobscura*. Using the estimates $E[I_2] = 15$, $D = 12{,}000$, and $P_1 = 0.1$, equations (1) and (14) together yield $\hat{K} \sim 10^6$. If this were the case, $D \gg \hat{K}$ (as required for eq. 14) would not hold for the more closely related comparisons used by Kondrashov et al. (2002), supporting the intuition that a saturating model could not account for their results. Clearly, however, the sources of error in the estimation of $\hat{K}$ are extremely numerous (Orr and Turelli 2001).

Having discussed a range of model based on the insights of Bateson (1909), Dobzhansky (1937), and Muller (1942), the final section briefly reviews a quite distinct class of models.

Wrightian Models

Bateson, Dobzhansky, and Muller showed that sets of complementary alleles could accumulate when populations evolved via a chain of fit intermediates. It is possible, however, for populations to pass between mutually incompatible states when no fit intermediates exist. In general, this requires a stochastic transition in which genetic drift is able to counteract selection. During this transition, the incompatibility may reach high frequency. Sewall Wright saw the occurrence of such stochastic transitions as an important part of adaptive evolution (e.g., Wright 1932), and as such, they formed a key component of his shifting balance theory (Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000).

However plausible we consider Wright's theory in general, two considerations argue against the importance of Wrightian processes in the accumulation of the intrinsic genetic incompatibilities that distinguish species. The first is the severity of the fitness loss caused by many individual incompatibilities. A variety of theoretical work has shown that when intermediates show greatly reduced fitness, stochastic transitions between the mutually incompatible states become ex-

tremely rare (e.g., Lande 1979; Walsh 1982; Coyne et al. 1997, 2000; Gavrilets 2003). The second consideration is the asymmetrical nature of incompatibilities. This, as has been shown, is a clear prediction of the models related to the work of Bateson, Dobzhansky, and Muller but not of Wrightian models.

Despite these basic difficulties, several authors have suggested that some broadly Wrightian process might help to account for some of the patterns in the incompatibility data (Johnson 2000; Gourbiere and Mallet, unpubl. ms; see also Palopoli and Wu 1994; Wade and Goodnight 1998; Goodnight and Wade 2000). With this in mind, three classes of Wrightian model are now considered. These deal with incompatibilities forming at, respectively, single sites, pairs of sites, and three or more sites.

Nei et al. (1983), Gavrilets and Hastings (1996), and Gourbiere and Mallet (unpubl. ms) analyzed models in which incompatibilities may form from a single substitution at a diploid locus. This requires the drift-aided fixation of underdominant mutations (in which the heterozygote shows reduced fitness). Identical dynamics describe the fixation of underdominant chromosomal rearrangements (Lande 1979; Walsh 1982). Clearly, such models can result in a constant P_1. However, they cannot account for incompatibilities involving hemizygous sex chromosomes, on which much of the introgression work has focussed (e.g., Orr and Coyne 1989), nor for the data used by Kondrashov et al. (2002), which includes recessive pathogenic mutations (e.g., Schaner et al. 2001).

Kimura (1985, 1991) introduced a model of compensatory neutral mutations occurring at pairs of sites. Under this model, mutants are assumed to be deleterious individually, but neutral when they occur in pairs. This model is closely related to the epistatic selection model, but the differences have important dynamical consequences. In one parameter regime, when the sites are closely linked, and both single mutants show a substantial reduction in fitness, then substitutions tend to occur only as simultaneous pairs (Kimura 1985; Stephan 1996). If these double substitutions constitute a constant fraction of the divergence, then a compensatory neutral model might account for Kondrashov et al.'s finding of a constant P_1 (the unique prediction being that all incompatibilities should be fully symmetrical). In an alternative parameter regime, when one of the single mutants is very mildly deleterious, we have a special case of Dobzhansky's (1937) two-locus model. In this case, we would lose the temporal coupling of substitution pairs that leads to the constant P_1 (Innan and Stephan 2001; Kondrashov et al. 2002).

In contrast to these one- and two-site Wrightian models, Johnson (2000; see also Palopoli and Wu 1994) suggested that the existence of complex incompatibilities, involving multiple factors (as evidenced by the introgression studies), might indicate that the populations had diverged via a broadly Wrightian process. It is not clear, however, that the assembly of complementary alleles by Wrightian means would lead to a situation where multiple, but not individual, introgressions led to an incompatibility. It seems equally likely that a single introgression could disrupt the coadaptation. In addition, Cabot et al. (1994) and Orr (1995) argued that divergence taking place via single sequential substitutions favors the formation of complex incompatibilities, that is, DMIs with high n (see above, and the Appendix). No such arguments apply to Wrightian models.

Discussion

This study has compared simple theoretical models of the evolution of intrinsic postzygotic reproductive isolation. There has been a focus on the ability of each model to explain three basic patterns in the empirical literature: (1) the asymmetrical nature of incompatibilities under reciprocal introgression; (2) the discovery of many incompatibilities requiring multiple concurrent introgressions; (3) and the constant probability, over a large range of divergence, of obtaining an incompatibility with a single introgression.

In addition, the study has asked whether any of these patterns might allow us to make inferences about the substitution process behind the divergence of the populations—whether patterns in the incompatibility data might be reliable signatures of particular modes of evolution. With this in mind, a distinction has been made between models that are agnostic about the evolutionary forces driving the substitutions and those that make strong assumptions about the mode of divergence. The latter category is further divided into models consistent with the ideas of Bateson, Dobzhansky, and Muller (in which the divergence may take place via sequential substitutions, each unopposed by natural selection), and Wrightian models (those requiring stochastic shifts between genetic combinations lacking fit intermediates). All of these approaches are equally principled. Of course, we may suspect that two sites capable of forming an incompatibility have some kind of functional interaction in normal development (although this does not have to be the case). But it is a quite different to propose that this interaction played an important role in the substitution process. Discriminating between the alternatives is an empirical question. How successful, then, are the various models in accounting for the empirical findings?

First, and most conclusively, the data support a model based on the work of Bateson, Dobzhansky, and Muller. This conclusion follows chiefly from pattern 1, the finding that incompatibilities tend to be asymmetrical (Wu and Beckenbach 1983; Orr and Coyne 1989; Coyne and Orr 1998). This is a consistent prediction of the Bateson-Dobzhansky-Muller models. In contrast, the Wrightian alternatives, because they assume that populations pass between mutually incompatible states, would seem to predict symmetrical incompatibilities. This conclusion is not contradicted by the other empirical patterns, which are in no case more plausibly explained by Wrightian evolution. Additional support comes from the theoretical finding that stochastic transitions will be extremely rare when the intermediates involve substantial fitness loss (e.g., Lande 1979; Gavrilets 2003).

To distinguish between the various Bateson-Dobzhansky-Muller models, the other patterns in the data must be considered. Consider next, pattern 2, the detection of DMIs requiring multiple concurrent introgressions. This pattern has been shown to be consistent with all of the models considered, with the exception of Kondrashov et al.'s (2002) epistatic selection model. A modified version of that model,

incorporating genetic redundancy, was introduced, and this could account for the findings. It is notable, however, that the modification of Kondrashov et al.'s model was rather ad hoc. Indeed, in no case did the prediction of pattern 2 follow from assumptions about the substitution process. As such, there is little chance of making inferences about the mode of divergence from the detection of these multiple-introgression DMIs.

A distinct question concerns the predictions of the models as to how readily such DMIs will be detected. Orr's combinatorial model, and even more so the epistatic selection model with genetic redundancy, predict that, when the divergence is substantial, the probability of obtaining a DMI will decrease very rapidly with m, the number of introgressed factors required for the DMI to form (see eqs. 6, 7, and 10). In contrast, models that allow for some saturation—meaning that the number of DMIs does not increase after a certain level of divergence—do not show this rapid decline (e.g., eq. 11). The substantial number of multiple-introgression DMIs that have been recorded (Coyne and Orr 1998; Johnson 2000) does suggest, at least naively, that they are fairly common. At present, however, no useful quantitative statement can be made, and so no clear conclusions can be drawn.

Pattern 3, the finding of a constant P_1, is the least well established (see above). It is also the most promising candidate as a signature of a particular mode of divergence. This is because the pattern is inconsistent with the prediction of Orr's (1995) combinatorial model, a model that predicts a snowballing of DMIs with genetic distance (eq. 5). Instead, the pattern is consistent with the prediction of the epistatic selection models, which rely on strong assumptions about divergence (eqs. 8, 9). However, it has been shown here that a constant P_1 is also consistent with models in which the number of DMIs saturates after a certain level of divergence (eqs. 11, 9). Because these saturating models do not make assumptions about the mode of divergence, this suggests that a constant P_1 may not be a reliable indicator of epistatic selection. It should be borne in mind, however, that explanations based on saturation could only apply at very large divergences. For this reason, it remains doubtful whether such an explanation can plausibly account for the results of Kondrashov et al. (2002). All of the above considerations suggest that studies similar to that of Kondrashov et al., but concentrating on closely related species pairs, would be of great importance. Such studies might have to take into account complications arising from divergence in parapatry (e.g., Gavrilets 1999; Kondrashov 2003).

Another quite distinct way of discriminating between the alternative models would be to directly investigate the relevant context of a particular substitution. In other words, we would like to know the number and position of sites that may influence the effects on fitness of a given substitution. (Similar ideas are often expressed in terms of the degree of correlation of a fitness landscape; Fontana et al. 1993; Gavrilets 1997.) The differing predictions of the Bateson-Dobzhansky-Muller models investigated here stem, in part, from their different assumptions in this respect. Biochemical modeling of individual molecules (e.g., Fontana et al. 1993; Kirby et al. 1995; Sunyaev et al. 2000; Bastolla et al. 2002; Kondrashov et al. 2002) and careful experimental manipulations (e.g., Elena and Lenski 2001; Ungerer et al. 2003) are making such questions increasingly tractable.

Acknowledgments

I am very grateful to J. Peck, D. Waxman, H. A. Orr, S. Gavrilets, A. Kondrashov, and I. Harvey who each provided detailed and extremely helpful comments on the manuscript. I would also like to thank M. Woolfit, T. Price, S. Gourbiere, and L. Bromham for helpful discussions. This work was supported by the University of Sussex under its Graduate Teaching Assistantship scheme.

Literature Cited

Barraclough, T. G., A. P. Vogler, and P. H. Harvey. 1998. Revealing the factors that promote speciation. Philos. Trans. R. Soc. B 353: 241–249.

Barton, N. H. 2001. The role of hybridization in evolution. Mol. Ecol. 10:551–568.

Bastolla, U., M. Porto, H. E. Roman, and M. Vendruscolo. 2002. Lack of self-averaging in neutral evolution of proteins. Phys. Rev. Lett. 89:208101.

Bateson, G. 1909. Heredity and variation in modern lights. Pp. 85–101 *in* A. C. Seward, ed. Darwin and modern science. Cambridge Univ. Press, Cambridge, U.K.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat. Rev. Genet. 4:216–224.

Cabot, E. L., A. W. Davis, N. A. Johnson, and C. -I. Wu. 1994. Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics 137:175–189.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

———. 1997. ''Patterns of speciation'' revisited. Evolution 51: 295–303.

———. 1998. The evolutionary genetics of speciation. Philos. Trans. R. Soc. Lond. B 353:287–305.

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. A critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643–671.

———. 2000. Is Wright's shifting balance process important in evolution? Evolution 54:306–317.

Davis A. W., E. G. Noonburg, and C.-I. Wu. 1994. Evidence for complex genic interactions between conspecific chromosomes underlying hybrid female sterility in the *Drosophila simulans* clade. Genetics 137:191–199.

Dobzhansky, T. G. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Edmands, S. 2002. Does parental divergence predict reproductive compatibility? Trends Ecol. Evol. 17:520–527.

Elena, S. F., and R. E. Lenski. 2001. Epistasis between new mutations and genetic background and a test of genetic canalization. Evolution 55:1746–1752.

Fay, J. C., G. J. Wyckoff, and C.-I. Wu. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature 415:1024–1026.

Fontana, W., P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. 1993. RNA folding and combinatory landscapes. Phys. Rev. E 47:2083–2099.

Gavrilets, S. 1997. Evolution and speciation on holey adaptive landscapes. Trends Ecol. Evol. 12:307–312.

———. 1999. A dynamical theory of speciation on holey adaptive landscapes. Am. Nat. 154:1–22.

———. 2003. Models of speciation: What have we learned in 40 years? Evolution 57:2197–2215.

Gavrilets. S., and J. Gravner 1997. Percolation of the fitness hypercube and the evolution of reproductive isolation. J. Theor. Biol. 184:51–64.

Gavrilets, S., and A. Hastings. 1996. Founder effect speciation: a theoretical reassessment. Am. Nat. 147:466–491.
Gillespie, J. H. 1984. Molecular evolution over the mutational landscape. Evolution 38:1116–1129.
———. 1991. The causes of molecular evolution. Oxford Univ. Press, New York.
Goodnight, C. J., and M. J. Wade. 2000. The ongoing synthesis: a reply to Coyne, Barton, and Turelli. Evolution 54:317–324.
Innan, H., and W. Stephan. 2001. Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. Genetics 159:389–399.
Johnson, N. A. 2000. Gene interactions and the origin of species. Pp. 197–212 *in* J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford Univ. Press, Oxford, U.K.
———. 2002. Sixty years after ''Isolating mechanisms, evolution and temperature'': Muller's legacy. Genetics 161:939–944.
Kimura, M. 1985. The role of compensatory neutral mutations in molecular evolution. J. Genetics 64:7–19.
———. 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. Proc. Natl. Acad. Sci. USA 88:5969–5973.
Kirby, D. A., S. V. Muse, and W. Stephan. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA 92:9047–9051.
Kondrashov, A. S. 2003. Accumulation of Dobzhansky-Muller incompatibilities with a spatially structured population. Evolution 57:151–153.
Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. USA 99:14878–14883.
Kondrashov, F. A., and A. S. Kondrashov. 2001. Multidimensional epistasis and the disadvantage of sex. Proc. Natl. Acad. Sci. USA 98:12089–12092.
Lande, R. 1979. Effective deme size during long-term evolution estimated from rates of chromosomal rearrangement. Evolution 33:234–251.
Lijtmaer, D. A., B. Mahler, and P. L. Tubaro. 2003. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution 57:1411–1418.
Maynard Smith, J. 1978. The evolution of sex. Cambridge Univ. Press, Cambridge, U.K.
Muller, H. J. 1942. Isolating mechanisms, evolution and temperature. Biol. Symp. 6:71–125.
Nei, M., T. Maruyama, and C.-I. Wu. 1983. Models of evolution of reproductive isolation. Genetics 103:557–579.
Nowak, M. A., M. C. Boerlijst, J. Cooke, and J. Maynard Smith. 1997. Evolution of genetic redundancy. Nature 388:167–171.
Ohta, T. 1997. Role of random genetic drift in the evolution of interactive systems. J. Mol. Evol. 44:S9–S14.
Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.
———. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics 144:1331–1335.
———. 1997. Haldane's rule. A. Rev. Ecol. Syst. 28:195–218.
———. 2001. The genetics of species differences. Trends Ecol. Evol. 16:343–350.
Orr, H. A., and J. A. Coyne. 1989. The genetics of post-zygotic isolation in the *Drosophila virilis* group. Genetics 121:527–537.
Orr, H. A., and S. Irving. 2001. Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. Genetics 158:1089–1100.
Orr, H. A., and L. H. Orr. 1996. Waiting for speciation: the effect of population subdivision on the time to speciation. Evolution 38:1251–1749.
Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55:1085–1094.
Palopoli, M. F., and C.-I Wu. 1994. Genetics of hybrid male sterility between *Drosophila* sibling species: a complex web of epistasis is revealed in interspecific studies. Genetics 138:329–341.
Perez, D. E., and C.-I. Wu. 1995. Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. Genetics 140:201–206.
Perez, D. E., C.-I. Wu, N. A. Johnson, and M.-L. Wu. 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus (Ods)*. Genetics 134:261–275.
Peters, A. D., and C. M. Lively. 2000. Epistasis and the maintenance of sex. Pp. 99–112 *in* J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford Univ. Press, Oxford, U.K.
Polymeropoulos, M. H., C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, and R. L. Nussbaum. 1997. Mutation in the synuclein gene identified in families with parkinson's disease. Science 276:2045–2047.
Presgraves, D. C. 2002. Patterns of post-zygotic isolation in Lepidoptera. Evolution 56:1168–1183.
Price, T. D., and M. M. Bouvier. 2002. The evolution of F_1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.
Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we learned in 40 years? Evolution 47: 1637–1653.
Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns in postzygotic isolation in frogs. Evolution 52:1811–1820.
Schaner, P., N. Richards, A. Wadhwa, I. Aksentijevich, D. Kastner, P. Tucker, and D. Gumucio. 2001. Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. Nat. Gen. 27:318–321.
Smith, N. G. C., and A. Eyre-Walker. 2002. Adaptive protein evolution in *Drosophila*. Nature 415:1022–1024.
Stephan, W. 1996. The rate of compensatory evolution. Genetics 144:419–426.
Sunyaev, S., V. Ramensky, and Bork P. 2000. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet. 16:198–200.
Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16:330–343.
Ungerer, M. C., C. R. Linder, and L. H. Rieseberg. 2003. Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. Genetics 163:277–286.
Wade, M. J., and C. J. Goodnight. 1998. The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. Evolution 52:1537–1553.
Walsh, J. B. 1982. Rate of accumulation of reproductive isolation by chromosome rearrangements. Am. Nat. 120:510–532.
Wilf, H. S. 1990. Generatingfunctionology. Academic Press, Boston, MA.
Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. 6th Int. Congr. Genet. 1: 356–366.
Wu, C.-I., and A. T. Beckenbach. 1983. Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *D. persimilis* and identification of hybrid sterility factors. Genetics 105:71–86.

Corresponding Editor: J. Hey

Appendix

In this appendix, equations (2) and (3) are derived. In addition, the ''possible evolutionary paths'' arguments of Cabot et al. (1994) and Orr (1995) are very briefly presented.

Consider first equation (2), the number of unproven combinations of n amino acids that may be created from two genetic backgrounds differing by D amino acids. First, we note that this number must be proportional to the number of groups of n amino acids that can be formed from the D amino acids at which the backgrounds differ; this is simply the binomial coefficient, $\binom{D}{n}$ (Orr 1995). Since we are considering a pair of sequences, a total of 2^n distinct combinations of amino acids may be formed from the amino acids in each group. Of these combinations, given the assumptions about divergence in

the text, $n + 1$ will have been proven to confer high fitness—those containing each of the n nonsynonymous substitutions distinguishing the sequences at the relevant sites, plus the ancestral sequence. Together, this yields equation (2). Although the assumption of no multiple substitutions at the same site is certainly questionable when one considers the large genetic distances analyzed by Kondrashov et al. (2002), their inclusion would not help to reconcile theory with data. This is because, in general, multiple substitutions make proven combinations less likely to be reconstituted, and this would simply increase $C(D, n)$. The same applies to simultaneous substitutions (but only if the substitution dynamics can be uncoupled from the formation of potential DMIs, as the epistatic selection model makes clear).

Consider now equation (3), the expected number of unproven combinations formed by the introgression of m amino acids, counting only combinations that include all m of the introgressions. Given the compulsory inclusion of the m, there are a total of $\binom{D-m}{n-m}$ ways of choosing the remaining $n - m$ amino acids that make up the combination. The probability that such a combination is unproven will be denoted $F(n,m)$ and is given by

$$F(n, m) = \left[1 - 1 \Big/ \binom{n}{m}\right]. \tag{A1}$$

This result follows from arguments of Orr's (1995) and is, in fact, his equation (11). Combining these, the expected number of unproven combinations is simply

$$C(D, n, m) = \binom{D - m}{n - m} F(n, m), \tag{A2}$$

which can be rearranged to yield equation (3). Note that $C(D, n, m)$, unlike $C(D, n)$ is the expectation of a random variable. This is because substitutions that occurred at different times during the divergence may participate in different numbers of unproven combinations, a fact that explains why equation (4) is given in approximate form. Note that equation (2) may also be derived from

$$C(D, n) = \sum_{m=1}^{n-1} \binom{D}{m} C(D, n, m), \tag{A3}$$

which makes intuitive sense given the way these quantities are defined.

Orr (1995) introduces equation (A1), in a quite different context. Given a pair of divergent genotypes capable of forming an n-fold DMI, $F(n, m)$ gives the fraction of the possible evolutionary ''paths,'' consisting of single substitutions, that connect the two genotypes without including the DMI as an intermediate step. Since equation (A1) increases with n, Cabot et al. (1994) and Orr (1995) argue that it may be important in explaining the results of introgression studies. Note that, although $F(n, m)$ appears in equation (A2), the reasoning is quite different. Equation (A2) incorporates the probablity of reconstructing a previously unproven combination given that a certain level of divergence has taken place. The evolutionary path argument concerns the likelihood that divergence will take place at these sites, given that a DMI may be formed between them (Orr 1995). Such considerations would certainly alter results such as equation (6) and equation (7), however, it is unlikely that they would alter the qualitative conclusion that single introgressions are more likely to form DMIs. Calculating the quantitative effect would require an explicit model of divergence (as is clear from, e.g., Stephan 1996).