doi:10.1111/j.1558-5646.2009.00844.x

ARE SPECIES REAL? THE SHAPE OF THE SPECIES BOUNDARY WITH EXPONENTIAL FAILURE, REINFORCEMENT, AND THE "MISSING SNOWBALL"

Sébastien Gourbière[1,2] and James Mallet[3,4]

[1]*UMR 5244 CNRS-EPHE-UPVD, Laboratoire de Biologie et d'Ecologie Tropicale et Méditerranéenne, Université de Perpignan, Via Domitia, 52 Avenue Paul Alduy, 66 860 Perpignan Cedex, France*

[2]*E-mail: gourbier@univ-perp.fr*

[3]*Wissenschaftskolleg zu Berlin, Wallotstraße 19, 14193 Berlin, Germany*

[4]*Galton Laboratory, Department of Genetics, Evolution and Environment, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom*

Received August 26, 2009
Accepted August 27, 2009

Under simple assumptions, the evolution of epistatic "Dobzhansky–Muller" incompatibilities between a pair of species should yield an accelerating decline of log overall reproductive compatibility—a "snowball" effect that might rapidly provide new species with "reality." Possible alternatives include: (1) simple exponential failure, giving a linear rate of log compatibility loss, and (2) "slowdown," likely during reinforcement in which mate choice evolves to prevent deleterious hybridization, yielding a decelerating log compatibility loss. In analyses of multiple datasets, we find little support for the snowball effect, except possibly in Lepidoptera hybrid viability. The snowball predicts a slow initial rate of incompatibility acquisition, with low initial variance; instead, highly variable compatibility is almost universally observed at low genetic distances. Another deviation from predictions is that reproductive isolation usually remains incomplete until long after speciation. These results do not disprove snowball compatibility decay, but can result if large deleterious effects are due to relatively few genetic changes, or if different types of incompatibility evolve at very different rates. On the other hand, data on *Bacillus* and *Saccharomyces*, as well as theories of chromosomal evolution, suggest that some kinds of incompatibility accumulate approximately linearly, without Dobzhansky–Muller effects. In microorganisms, linearity can result from direct negative effects of DNA sequence divergence on compatibility. Finally, a decelerating slowdown model is supported for sympatric *Leptasterias* starfish, and in *Drosophila* prezygotic isolation in sympatry but not allopatry, providing novel comparative evidence for reinforcement.

KEY WORDS: Chromosomal incompatibility, comparative analysis, Dobzhansky–Muller incompatibilities, reproductive isolation, reinforcement.

It has been argued that reproductive isolation provides a good definition of species because the emerging "reality" of species is ensured by barriers to gene flow. This would especially be the case if overall barriers caused by postzygotic isolation rapidly accelerated after the initiation of speciation, effectively shutting the door to interspecific gene flow. In this article, we collate information on many types of reproductive isolation and its effect on sexual compatibility to investigate the time course of the evolution of barriers to gene flow.

Compatibility among sexual taxa declines as a result of two rather distinct processes during or after speciation, both resulting in a state referred to as reproductive isolation. First, mating

behavior or gametic recognition may diverge and reduce the rate of interpopulation fertilization, to cause "prezygotic isolation." Second, genetic changes between taxa may cause hybrid sterility or inviability, or "postzygotic isolation." The processes are very different: prezygotic isolation directly reduces the level of gene flow; postzygotic isolation selects against the genes that have flowed, altering only "effective" or "successful" gene flow, rather than actual gene flow. Prezygotic isolation may be directly selected to avoid gametic wastage and unfit offspring via a good-genes sexual selection process known as "reinforcement," whereas postzygotic isolation will usually evolve as an indirect, pleiotropic result of genetic divergence (the reasons for this divergence itself may be either neutral drift, or selection for traits other than reproductive isolation).

In recent years, evolutionary biologists have considerably advanced understanding of reproductive isolation. Recent work has included: (1) detailed genomic mapping of particular pairs of species, leading to an understanding of the numbers and genetic architecture of genes causing pre- and postzygotic isolation (Hollocher and Wu 1996; Rieseberg et al. 1996; True et al. 1996; Wu et al. 1996; Dopman et al. 2004); (2) studies of individual "speciation genes," mainly in *Drosophila* (Orr et al. 2004); (3) theoretical analyses of hybrid inviability and sterility (Orr and Turelli 2001; Gavrilets 2003; Welch 2004; Gavrilets 2004; Turelli and Moyle 2007), particularly Haldane's rule, the tendency for the heterogametic sex of hybrids to suffer more inviability and sterility than the homogametic sex (Turelli and Orr 1995, 2000; Turelli and Moyle 2007); (4) theoretical studies of the evolution of ecological divergence, mate recognition and reinforcement, the tendency for pairs of taxa to diverge ecologically and in mate recognition in the face of gene flow (Kirkpatrick and Ravigné 2002; Servedio and Noor 2003); and (5) comparative analyses of the evolution of reproductive isolation related to time of divergence (Coyne and Orr 1997; Sasa et al. 1998; Fitzpatrick 2002; Presgraves 2002; Price and Bouvier 2002; Mendelson et al. 2004; Bolnick and Near 2005). In spite of this work, many problems remain. For example, some argue that ecological divergence and prezygotic isolation is most likely to drive speciation, with postzygotic isolation becoming important only after speciation is more or less complete (Mallet et al. 1998; Price and Bouvier 2002; Jiggins et al. 2005; Mallet 2006). Others argue that prezygotic and environment-independent postzygotic isolation are about equally important, because in allopatric taxa of *Drosophila* they evolve at approximately the same rate (Wu 1996; Coyne and Orr 1997, 2004), or because hybridization still occurs in nature between species with strong postzygotic isolation (Presgraves 2002).

The theory of "Dobzhansky–Muller incompatibilities" has suggested that epistatic postzygotic incompatibilities should evolve as a kind of "snowball" effect, i.e., with increasing rapidity as populations diverge (Orr 1995; Orr and Turelli 2001). Incompatibilities depending on exactly two substitutions are expected to be established as the square of the time since divergence, and incompatibilities depending on more substitutions are established with correspondingly greater acceleration. The idea that low fitness of hybrids is mainly due to interactions (epistasis) among two or more loci, as opposed to heterozygote disadvantage at single loci, is extremely compelling: such epistasis can explain well-known examples of postzygotic isolation, such as Haldane's rule (Gavrilets 2003; Coyne and Orr 2004; Welch 2004). However, there has been little progress in formulating predictions of snowball theory that can be tested using comparative data (but see Kondrashov et al. 2002; Mendelson et al. 2004; Welch 2004; Bolnick and Near 2005; Turelli and Moyle 2007).

There are, moreover, several nonepistatic alternatives for the evolution of postzygotic isolation. For example, chromosomal rearrangements can lead simply to heterozygote disadvantage (Walsh 1982; Coyne and Orr 2004; Welch 2004; Gavrilets 2004; Kirkpatrick and Barton 2007). Incompatibilities due to chromosomal rearrangements appear to cause direct major deleterious hybrid fitness effects in mammals (Chandley 1988; Britton-Davidian et al. 2000), although they have been argued to be unimportant in *Drosophila* (Coyne et al. 1991). In yeast, elegant chromosomal engineering has shown that rearrangements themselves have direct effects on hybrid fertility, and do not just trap genes with epistatic fitness effects (Delneri et al. 2003; Greig 2007). Whether local adaptation (Kirkpatrick and Barton 2007) or genetic drift (Walsh 1982; Coyne et al. 1997; Gavrilets 2004) is the cause, nonepistatic selection of this type means that chromosomal rearrangements should accumulate approximately linearly with time. Nonepistatic hybrid unfitness could also be due to a deleterious effect of DNA divergence on recombination, which causes reduced fertility in yeast and reduced transformation efficiency in bacteria such as *Bacillus* (Zawadzki et al. 1995; Greig et al. 2003; Liti et al. 2006). As well as chromosomal evolution and direct effects of DNA divergence, divergence in quantitative traits may lead to ecologically based isolation involving little hybrid incompatibility. For example, some sorts of ecologically based assortative mating may be due to simple habitat divergence; hybrids may be able to use intermediate or both habitats (Drès and Mallet 2002; Jiggins et al. 2005). Because such kinds of incompatibility are simple (rather than more complex incompatibilities depending on two or more mutations, as in Dobzhansky–Muller epistasis), their accumulation should not deviate strongly from a linear model. Recent hybrid incompatibility theory has focused on epistatic models, particularly Dobzhansky–Muller incompatibilities, but we still have little idea of how epistatic and nonepistatic processes might interact to produce overall hybrid incompatibility (Kirkpatrick and Barton 2007; Turelli and Moyle 2007).

Prezygotic isolation could behave differently again. For example there might be diminishing returns if mating behavior is under selection to avoid the production of unfit hybrid offspring, i.e., reinforcement. Reinforcement is an old idea (Dobzhansky 1940), but recent work has confirmed its existence (Butlin 1995; Noor 1995; Higgie et al. 2000). In a pair of species successfully undergoing reinforcement, initial substitution of mutations causing assortative mating will be strongly favored. Once assortative mating is well established, there will be less and less selection for further assortative mating, because the selection pressure depends directly on the production of hybrids and level of gene flow. The substitution rate at loci affecting assortative mating should therefore decline during the process of reinforcement, giving a less-than-first-order rate process, a "slowdown" pattern. Although we have framed this idea in terms of sexual compatibility, a variety of multilocus adaptive processes evolving toward fixed optima can create conditions in which early phenotypic changes are more strongly selected, so that mutations for larger effect tend to be fixed, than in later stages in attaining the optimum (Orr 1998). A slowdown substitution process has already been suggested as an effect that might counteract the snowball effect on overall compatibility. Postzygotic isolation could then accumulate closer to linearly with time than under the snowball process alone (Mendelson et al. 2004). However, there seems no a priori reason why this should be so. In any case, slowdown processes may be a more general feature of adaptive evolution than hitherto realized.

In this article, we provide general models to cover a variety of possible modes of overall compatibility decline: less-than-first-order (slowdown), first-order (linear), and higher-order rate (snowball) models of incompatibility acquisition. Under simple assumptions, we argue that these phenomena may be produced as a result of epistatic, null-model exponential, and reinforcement processes, respectively. However, we emphasize that we are more interested in testing for evidence of curvilinearity on a log scale, that is for deviation from a simple exponential model, than in whether such curvilinearity is evidence for underlying processes. We then evaluate model predictions against comparative data from a variety of taxa via exploratory analyses of comparative data, and discuss the fit to the various models. Our curve-fitting approaches complement and go beyond other recent attempts to achieve similar ends: (1) We analyze a greater variety of data; (2) we develop for the first time a suitable method to handle discrete viability and fertility data (discretized data have become traditional in *Drosophila*, bird, frog, and Lepidoptera data (Coyne and Orr 1997; Presgraves 2002; Price and Bouvier 2002)); (3) our hypothesis-testing approach integrates appropriate theory with a simple multiplicative fitness function for incompatibilities; and (4) we develop and test the slowdown prediction for reinforcement for the first time.

Theory of Compatibility Decline

MODELLING THE TIME COURSE OF INCOMPATIBILITY EVOLUTION

We investigate several combinations of substitution processes and incompatibility interaction: (1) constant substitution assuming simple nonepistatic incompatibility (linear model); (2) decelerating substitution with nonepistatic incompatibility (slowdown model); and (3) constant substitution with epistatic incompatibility (snowball model). We do not test a more complex situation—variable substitution with epistatic incompatibilities, as suggested by Mendelson et al. (2004)——because an acceptable array of possible time-courses of incompatibility acquisition is already provided by the first three (Fig. 1). Here, we justify the three alternative models microscopically in terms of two-locus epistasis, substitution rate variation, and multiple effects on fitness.

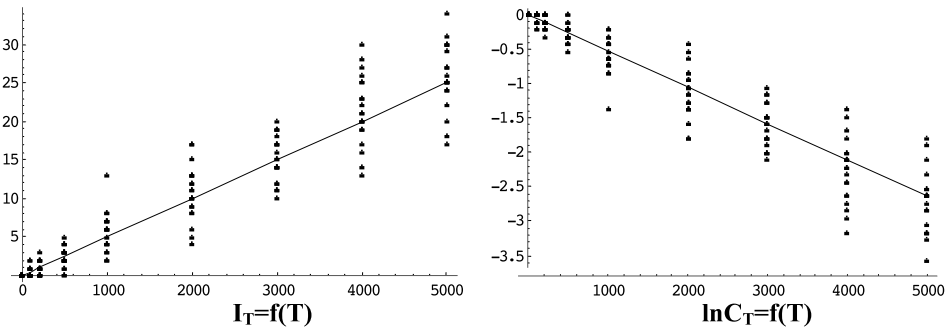
In the simplest scenario, we consider a constant substitution rate and nonepistatic incompatibilities each caused by a single substitution. These assumptions lead to a linear rate of accumulation of incompatibilities. If individual incompatibilities are small and independent, they will combine multiplicatively (Orr 1995), and overall compatibility or hybrid fitness will decline approximately exponentially (Walsh 1982; Gavrilets 2004). This exponential distribution is typical of the failure of many mechanical or electrical components (e.g., light bulbs), or life span in organisms such as invertebrates that do not show ageing. It is known as the "exponential failure law" in engineering, or the "type I survival curve" in population dynamics. In some ways, this simple linear or first-order model can be seen as the simplest null hypothesis against which more complex nonlinear models must be tested.

A second scenario accounts for a constant substitution rate and two-locus Dobzhansky–Muller incompatibilities. As demonstrated by Orr and Turelli, with these two assumptions the number of incompatibilities increases (snowballs) with the square of time since divergence (Orr 1995; Orr and Turelli 2001). The third scenario includes both variable substitution rates and nonepistatic incompatibilities. Because we are primarily interested in modeling a decreasing substitution rate (see introduction), the acquisition of incompatibilities will be less than linear with time, the slowdown model.

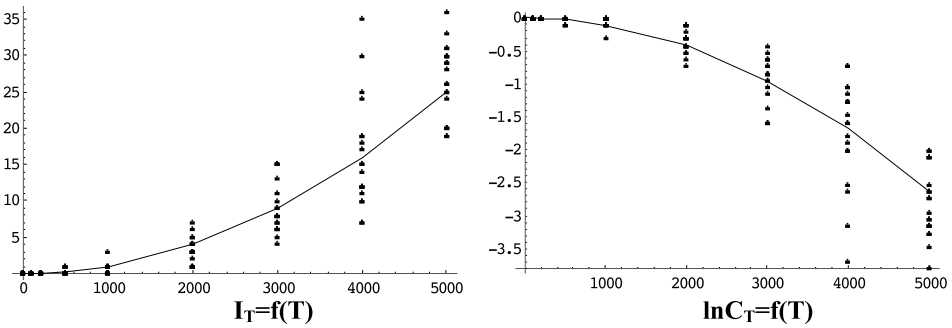
VARIABLE SUBSTITUTION RATE AND THE "SLOWDOWN" MODEL

A simple assumption in molecular evolution is that the molecular clock ticks at a constant rate if the changes are neutral. Given that substitution events are independent of one another, the number of substitutions between two lineages that diverged T generations ago follows a Poisson distribution with mean $\lambda = 2kT$, where k is the constant substitution rate in each lineage. This substitution process is the basis of the snowball model (Orr and Turelli 2001), as well as of linear models (Walsh 1982; Gavrilets 2004). A

A Linear. Constant substitution rate and single gene incompatibilities (p=0.1, s=0.2, and K= 0.025).



B Snowball. Constant substitution rate and Dobzhansky-Muller incompatibilities (p=0.005, s=0.1, and K= 0.01). Note especially the low initial variance.



C Slowdown. Decreasing substitution rate and single gene incompatibilities (p=0.1, s=0.1, and K= 0.35, a=0.01).

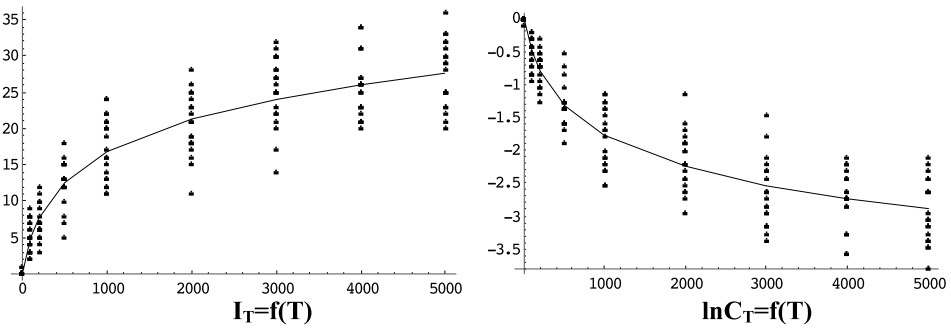


Figure 1. Relationship between numbers of incompatibilities (I_T) and log overall compatibility (ln C_T, on the ordinate) and the time of divergence (T, on the abscissa). Expectations and simulated data.

generalization of this substitution law will be able to model a range of variable substitution rates corresponding to any real non-negative function of time, $S(t)$. The number of substitutions K_T separating two taxa that diverged T years ago is then distributed according to

$$p(K_T = n) = (W_T)^n e^{-WT}/n! \quad \text{where } W_T = \int_0^T S(t')\,dt'. \quad (1)$$

This expression simplifies to the standard Poisson distribution when the substitution rate is constant, i.e., when $S(t) = 2k$ and $W_T = 2kT$. In the following, we will consider a more complex $S(t)$ function than $S(t) = 2k$, to allow a continuum between constant and variable substitution rates. We set $S(t) = 2k/(1 + at)$. This was chosen to be a simple algebraic form capable of producing a decrease in substitution rate (the higher the value of a, the faster the deceleration in the substitution rate), and the concave curvature

we require. When $a = 0$, $S(t) = 2k$ and the model reverts to a constant substitution rate (i.e., the linear model). Values of $a < 0$ correspond to a snowball-like acceleration of substitution rate, although the model does not deal with this (snowball-like) region very gracefully, as it develops an undefined substitution rate when $a < -t^{-1}$.

ACCUMULATION OF INCOMPATIBILITIES

We define the number I_T of incompatibilities after T years of divergence from the number of substitutions K_T according to the two kinds of incompatibilities we consider, i.e., according to a single mutation (linear) incompatibility scheme or an epistatic Dobzhansky–Muller (snowball) scheme. Following Orr and Turelli (2001), we assume that any pair of diverged sites suffers a small probability, P, of causing an incompatibility.

COMPATIBILITY BETWEEN SPECIES

Reproductive isolation consists of two components: assortative fertilization (normally considered "reproductive isolation" only if it occurs between species; when it occurs within species, assortative mating is considered to lead instead to "sexual selection"), and natural selection against deleterious genotypes (again normally only considered a form of reproductive isolation if it occurs between species). It should be noted that much, indeed perhaps most, "reproductive isolation" evolves long after speciation is generally accepted to have taken place, even though many evolutionary biologists claim to be using a reproductive isolation concept of species. The centrarchid fish form perhaps the most extreme example. In this group, sister species are often around 2 million years old, but natural hybridization does not cease until around 16 million years, and postzygotic reproductive isolation only becomes more or less complete by about 30 million years after the initial divergence (Bolnick and Near 2005). Here, we frame our analysis in terms of fitness, or progressive failure of compatibility, and we treat prezygotic and postzygotic components separately as far as possible. Compatibility is 100% if there is no assortative mating or selection against hybrids, and 0% if assortative mating or hybrid inviability and/or sterility is complete, and is equivalent to hybridization rate for prezygotic isolation, and to hybrid fitness for postzygotic data.

We map theories for the accumulation of incompatibilities onto an overall fitness or "reproductive isolation" scale via a simple multiplicative model whereby different incompatibilities combine to affect fitness. Every incompatibility is assumed to have an identical deleterious effect, s, on the fitness of hybrids (Walsh 1982; Gavrilets 2004). Provided that incompatibility effects are small and not too variable, the assumption has little effect on the results of this kind of model (Orr and Turelli 2001). If incompatibilities have large and highly variable effects, it is hard to fit any model (see discussion). We implement multiplicative fitness to predict compatibility between species that diverged T generations ago by setting $C_T = (1 - s)^{I_T}$ where s is the decrease in compatibility due to a single incompatibility (which itself may be due to a single mutation, or to an epistatic effect of several loci), and I_T is the number of incompatibilities at time T (see Fig. 1).

Previously, the problem of how incompatibilities are combined to affect fitness seems usually to have been more or less ignored (Fitzpatrick 2002; Mendelson et al. 2004; Bolnick and Near 2005), or was treated in a framework that assumes the number of incompatibilities evolves toward a threshold value representing complete reproductive isolation, viewed as equivalent to speciation (Orr and Turelli 2001). More recently, the latter framework has been made more flexible by allowing various curvilinear responses of hybrid fitness to the numbers of incompatibilities already existing (Turelli and Moyle 2007). These authors were mainly interested in modeling numbers of incompatibilities rather than the overall strength of reproductive isolation. However, some assumptions about combining incompatibilities are required to model the effects on overall compatibility (see note 1 below).

We use a compatibility failure approach, which is more standard in other types of survival analysis, for example in models of mechanical failure or life spans of organisms in populations. Multiplicative fitness combination seems most logical based on existing data for deleterious mutations (Charlesworth et al. 2004), and on classical probability theory of independent events as normally used in population genetics (Orr 1995; Mendelson et al. 2004). Fitness are multiplicative if the survival probability of a zygote with an incompatibility later in development or postpartum life is unaffected by the number of earlier challenges survived. Multiplicative fitness models the decline of survival as an asymptote to zero survival (Gavrilets 2004), rather than letting survival reach zero, as in threshold models (Orr and Turelli 2001). Although this may seem unrealistic for speciation, it is in fact sensible. Even though reasonably large-scale experiments may show apparently "complete" reproductive isolation, extremely rare combinations of events can allow occasional breeding success. For example, female mules and hinnies (horse × donkey hybrids) very occasionally produce viable backcross foals, although they were until recently viewed by geneticists as completely sterile. In this case, very rare coincidences combine so that chromosomal segregation in hybrid meiosis can occasionally produce viable gametes with complete genomic complements (Chandley 1988).

For each of the three scenarii, we now obtain the expectation and variance of the number of substitutions (K_T), the number of incompatibilities (I_T), and the compatibility between species (C_T) after T years of divergence (Appendix A). Table 1 summarizes the theoretical results used to fit the three models to the datasets described in the following section. Figure 1 shows simulated results in terms of the numbers of incompatibilities (I_T, K_T) and overall hybrid fitness (C_T).

Table 1. Expectation and variance of (1) the number of substitutions K_T, (2) the number of incompatibilities I_T, and (3) the compatibility after T years of divergence C_T between two species. The compatibility is defined as $C_T=(1-s)^{I_T}$, where s is the deleterious effect of any single incompatibility. For derivations, see Appendix A. Expectation and variance of $\ln C_T$ are linked to the expectation and variance of I_T in simple ways that do not depend on the model being considered. To obtain the actual expression of these two moments (not given to save space) one simply has to substitute $E(I_T)$ and $V(I_T)$ with their expression as functions of parameters of the model being considered.

	Linear model	Snowball model	Slowdown model
Expectation of K_T: $E(K_T)$	$2kT$	$2kT$	$[2k\ln(1+aT)]/a$, $a>-T^{-1}$
Variance of K_T: $V(K_T)$	$2kT$	$2kT$	$[2k\ln(1+aT)]/a$, $a>-T^{-1}$
Expectation of I_T: $E(I_T)$	$2kpT$	$2k^2T^2p$	$[2kp\ln(1+aT)]/a$, $a>-T^{-1}$
Variance of I_T: $V(I_T)$	$2ktpT$	$2k^2T^2p(1+4pkT)$	$[2kp\ln(1+aT)]/a$, $a>-T^{-1}$
Expectation of $\ln C_T$: $E(\ln C_T)$		$\ln(1-s)\,E(I_T)$	
Variance of $\ln C_T$: $\mathrm{Var}(\ln C_T)$		$\ln^2(1-s)\,V(I_T)$	

Datasets

We obtained data from a variety of publications listing measures of genetic distance, as well as isolation or compatibility of microbes, prezygotic isolation (between heterospecific males and females), and/or postzygotic isolation (inviability, sterility of F_1 hybrids). We split these datasets into prezygotic datasets (in which are included microbial datasets, and all those involving some degree of prezygotic isolation; see Tables 2 and 4), and postzygotic datasets (Tables 3 and 5). Some of these have already been used in a different way in earlier meta-analyses (Fitzpatrick 2002; Mendelson et al. 2004). Detailed notes on individual datasets are given in Appendix B.

Fitting

According to the models introduced above, the fits to the data are given by expectations on a linear compatibility scale as follows:

$$\text{Linear: } E(C_T) = \mathrm{Exp}[-e_1 T] \quad \text{where } e_1 = 2kps, \tag{2}$$

$$\text{Snowball: } E(C_T) = \mathrm{Exp}[-e_2 T^2] \quad \text{where } e_2 = 2k^2ps, \tag{3}$$

$$\text{Slowdown: } E(C_T) = \mathrm{Exp}[-e_3 \ln[(1+aT)]/a] \quad \text{where } e_3 = 2kps. \tag{4}$$

Alternatively, we can fit log-transformed compatibility data as follows:

$$\text{Linear: } E(\ln C_T) = e_1 T \quad \text{where } e_1 = 2kp\ln(1-s), \tag{5}$$

$$\text{Snowball: } E(\ln C_T) = e_2 T^2 \quad \text{where } e_2 = 2k^2 p\ln(1-s), \tag{6}$$

$$\text{Slowdown: } E(\ln C_T) = e_3 \ln[(1+aT)]/a \quad \text{where } e_3 = 2kp\ln(1-s). \tag{7}$$

We here use the term "snowball" in a general sense to refer to the decline of compatibility overall[1] as in Orr (1995), rather than as in Orr and Turelli (2001). Optimal fits of these models to datasets were obtained by a least squares approach. We performed all the fits reported here by minimizing unweighted sums of square deviations from expectations of compatibility. Because the models are inherently heteroscedastic, we also attempted weighted log-transformed fits by dividing sums of squares by the variance expected of log C_T (Table 1; however, we were unable to find similar expressions for variances of untransformed C_T). As conclusions were similar whether weighted or unweighted sums of squares were used, we report only unweighted analyses. There are a number of other statistical problems with comparative data meta-analyses of this kind, but we are more interested here in exploratory data analysis (Mosteller and Tukey 1977) of the shape of the decline in compatibility, rather than in testing the null hypothesis that no relationship of any kind exists between compatibility and genetic distance. A discussion of some of the many statistical problems is given in Appendix C.

CONTINUOUS COMPATIBILITY DATA

For all datasets, expected compatibilities are given by equations (2–7). Analytical expressions for least square estimators of e_1 and e_2 can then be obtained. Considering the linear model, the value of e_1 that minimizes the sum of squares is

[1] M. Turelli (pers. comm.) contends that our "snowball" model of overall compatibility decline is not necessarily an outcome of Orr's snowball model of incompatibility number accumulation, because virtually nothing is known about the way in which different incompatibilities combine in nature. However, it seems clear that the "snowball" model as originally formulated was intended to extend to overall compatibility in this way, rather than just to numbers of incompatibilities of unspecified overall effect (Orr 1995). Overall compatibility can, we argue, "snowball," just as can numbers of incompatibilities, and it is the former we test here. We will typically, in speciation research, be more interested in overall compatibility (or overall reproductive isolation) than in the numbers of incompatibilities, and this is the approach we adopt here.

Table 2. Prezygotic compatibility (fitted on untransformed data).

Model fitted	Sum of squares	Wilcoxon–Mann–Whitney P	Runs test P	e_1, e_2 or e_3	a
Bacillus, sexual compatibility vs. % divergence, N=53					
Snowball	0.98321	10^{-5}	2×10^{-4}	0.0572	
Linear	0.94179	4×10^{-4}	0.0072	0.266	
Slowdown	0.8622	0.381	0.0100	0.210	−0.070
Saccharomyces, spore viability vs. % divergence (JC-corrected), N=19					
Snowball	0.503	0.0002	0.0100	0.0440	
Linear	0.274	0.0420	0.544	0.218	
Slowdown	0.256	0.277	0.637	0.370	0.327
Leptasterias starfish, frequency of hybrids vs. % mtDNA div., N=9					
Snowball	0.00235	0.286	1	5964.194	
Linear	0.00133	0.286	1	131.306	
Slowdown	8×10^{-4}	0.286	1	246.960	89.220
Alpheus shrimps, behavioral compatibility vs. Nei's D, N=11					
Snowball	0.417	0.927	0.667	80.878	
Linear	0.644	0.042	0.285	9.584	
Slowdown	0.582	0.662	0.524	7.940	−3.676
Alpheus shrimps, behavioral compatibility vs. % mtDNA div., N=11					
Snowball	1.078	0.247	0.524	0.0110	
Linear	1.236	0.0727	0.285	0.108	
Slowdown	1.139	0.329	0.524	0.0714	−0.0520
Drosophila (sympatric), mating compatibility vs. Nei's D, N=45					
Snowball	1.607	0.147	0.541	3559.040	
Linear	1.668	0.426	0.968	59.354	
Slowdown	1.170	0.654	0.664	222.554	520.372
Drosophila (allopatric), mating compatibility vs. Nei's D, N=46					
Snowball	2.611	6×10^{-6}	3×10^{-4}	9.346	
Linear	1.907	0.492	0.869	2.601	
Slowdown	1.849	0.0186	0.734	3.348	1.282

$$e_1 = \frac{\sum_{i=1}^{n} t_i \ln C_i}{\sum_{i=1}^{n} t_i^2}. \quad (8)$$

In the snowball model, the value of e_2 that minimizes the sum of squares is

$$e_2 = \frac{\sum_{i=1}^{n} t_i^2 \ln C_i}{\sum_{i=1}^{n} t_i^4}. \quad (9)$$

We used these expressions to estimate e_1 and e_2. The estimates of e_3, a, and all sums of squares were evaluated numerically.

DISCRETE COMPATIBILITY DATA

In the data we use, postzygotic compatibility is often recorded as a set of discrete values, c_i. For example, in the *Drosophila* postzygotic data (Coyne and Orr 1997), the authors took the view that comparisons across heterogeneous datasets were simpler if only complete sterility and inviability of F_1 hybrids were recorded in each sex of F_1 hybrid and in each direction of cross. For example, if there were Haldane's rule (single sex) sterility in hybrids between A female × B male, while the reciprocal B female × A male cross produces some fertile males and females, the overall

Table 3. Postzygotic compatibility (fitted using untransformed data).

Model fitted	Sum of squares	Kruskall–Wallis *P*	Wilcoxon–Mann–Whitney *P*			e_1, e_2, or e_3	*a*
			(0,+)	(0,−)	(+,−)		
Drosophila (sympatric), discrete compatibility index, Nei's *D*, *N*=35							
Snowball	1.875	0.147	0.358	0.0420	0.0440	13.002	
Linear	1.875	0.906	0.490	0.341	0.387	3.026	
Slowdown	1.750	0.555	0.403	0.150	0.307	3.177	0.441
Drosophila (allopatric), discrete compatibility index, Nei's *D*, *N*=34							
Snowball	3.006	0.001	0.222	6×10^{-5}	0.002	3.748	
Linear	2.438	0.098	0.444	0.019	0.039	3.150	
Slowdown	2.438	0.098	0.444	0.019	0.039	3.190	0.061
Drosophila (combined), discrete compatibility index, Nei's *D*, *N*=69							
Snowball	5.625	0.001	0.310	2×10^{-4}	4×10^{-4}	10.660	
Linear	4.375	0.152	0.411	0.039	0.049	3.136	
Slowdown	4.250	0.016	0.868	2×10^{-4}	0.250	2.840	0.43
Lepidoptera, discrete viability index, Nei's *D*, *N*=69							
Snowball	2.062	0.0297	0.0230	0.0250	0.159	1.100	
Linear	2.687	4×10^{-5}	2×10^{-6}	0.0050	0.090	0.860	
Slowdown	2.687	4×10^{-5}	2×10^{-6}	0.0050	0.090	0.870	0.071
Lepidoptera, discrete overall compatibility, Nei's *D*, *N*=51							
Snowball	5.000	0.148	0.098	0.180	0.0270	13.312	
Linear	5.000	0.992	0.464	0.455	0.489	3.348	
Slowdown	5.000	0.699	0.272	0.220	0.406	3.870	0.18
Frogs, compatibility (EH)×(MET), Nei's *D*, *N*=89							
Snowball	13.131	0.203[1]	–	–	0.829[2]	48.768	
Linear	10.136	0.862[1]	–	–	0.932[2]	1.884	
Slowdown	8.004	0.174[1]	–	–	0.415[2]	57.440	198.001
Frogs, discrete compatibility index (1−*IPO*2), Nei's *D*, *N*=139							
Snowball	20.499	10^{-8}	0.0029	10^{-10}	9×10^{-8}	11.517	
Linear	20.499	10^{-7}	0.0340	10^{-12}	8×10^{-6}	2.748	
Slowdown	20.250	5×10^{-7}	0.0027	9×10^{-10}	8×10^{-7}	14.040	7.660
Birds, discrete compatibility index, $\Delta T_{50}H$, *N*=132							
Snowball	8.437	0.007	0.002	0.103	0.015	0.046	
Linear	7.516	0.776	0.260	0.388	0.324	0.200	
Slowdown	7.516	0.776	0.260	0.388	0.324	0.206	0.071
Birds, discrete compatibility index, γ-*HKY* corr. % *cytB* divergence, *N*=108							
Snowball	5.629	6×10^{-8}	10^{-7}	3×10^{-6}	0.034	0.005	
Linear	5.523	4×10^{-5}	10^{-5}	9×10^{-6}	0.281	0.066	
Slowdown	5.394	4×10^{-3}	0.339	0.009	5×10^{-4}	0.090	0.230

[1]Wilcoxon–Mann–Whitney test is used instead of a Kruskall–Wallis as there are only two categories of residuals.

[2]Runs test is used instead of multiple comparison test as there are only two categories of residuals.

Table 4. Prezygotic compatibility versus genetic distance (fitted using log-transformed data).

Model fitted	Sum of squares	Wilcoxon–Mann–Whitney P	Runs test P	e_1, e_2 or e_3	a
Bacillus, sexual compatibility vs. % divergence, N=53					
Snowball	48.199	0.160	0.0510	−0.038	
Linear	31.330	0.0485	0.0590	−0.401	
Slowdown	28.329	0.330	0.215	−0.329	-3.32×10^{-2}
Saccharomyces, spore viability vs. % divergence (JC-corrected), N=19					
Snowball	2.317	0.105	0.105	−0.027	
Linear	3.449	0.043	4×10^{-4}	−0.304	
Slowdown	0.953	0.737	0.497	−0.179	−0.059
Leptasterias starfish, frequency of hybrids vs. % mtDNA div., N=9					
Snowball	40.869	0.056	0.111	−1514.701	
Linear	9.115	0.111	0.429	−103.478	
Slowdown	5.936	0.905	0.429	−169. 861	25.567
Alpheus shrimps, behavioral compatibility vs. Nei's D, N=11					
Snowball	6.160	0.662	0.524	−70.844	
Linear	6.169	0.126	0.177	−14.488	
Slowdown	5.002	0.662	0.524	−10.207	−2.546
Alpheus shrimps, behavioral compatibility vs. % mtDNA div., N=11					
Snowball	12.848	0.247	0.524	−0.011	
Linear	11.926	0.788	0.285	−0.178	
Slowdown	11.176	0.788	0.285	−0.118	-3.79×10^{-2}
Drosophila (sympatric), mating compatibility vs. Nei's D, N=45					
Snowball	458.473	3×10^{-4}	0.005	−4.246	
Linear	287.774	10^{-4}	0.081	−5.259	
Slowdown	173.304	0.371	0.833	−156.03	192.712
Drosophila (allopatric), mating compatibility vs. Nei's D, N=46					
Snowball	382.337	2×10^{-5}	0.004	−1.753	
Linear	293.056	0.455	0.146	−3.114	
Slowdown	269.928	0.407	0.640	−8.421	3.142

compatibility was scored as 0.75. For such data, expected compatibilities must be computed from equations (2–7) in a different manner than for continuous data, and analytical results could not be obtained. All such fits were done numerically as follows (Fig. 2).

Consider a dataset including five possible discrete values of postzygotic isolation, denoted $c_1 = 1$, $c_2 = 0.75$, $c_3 = 0.5$, $c_4 = 0.25$, $c_{n=5} = 0$ (e.g., *Drosophila* postzygotic data; see Tables 3 and 5). For a particular set of parameter values of the model being considered, we evaluated from equations (2–7) the time t_j required for expected compatibility to reach exactly c_{j+1}. Expected compatibilities were then set to c_j when the observed time (genetic distance) is such that $t_{j-1} < t < t_j$, where $t_0 = 0$. Finally, compatibility is expected to be at the lowest value of compatibility, c_n when the observed genetic distance is larger than t_{n-1}. After evaluating the expected values of c_j, sums of squares were evaluated for each set of parameter value. We repeated this routine until parameter estimates changed by less than 10^{-3}.

FITTING $C = 0$ IN COMPATIBILITY DATA AFTER LOG-TRANSFORMATION

The lowest compatibility level c_n is usually zero. We used the data directly in untransformed compatibility fits, but we rescaled c_n for log compatibility fits for two reasons. First, it is impossible to fit an observed compatibility of $C = 0$ on a log compatibility scale. (Note, this problem does not affect analyses of untransformed

Table 5. Postzygotic compatibility (fitted using log-transformed data).

Model fitted	Sum of squares	Kruskall–Wallis *P*	Wilcoxon–Mann–Whitney *P*			e_1, e_2, or e_3	a
			(0,+)	(0,−)	(+,−)		
Drosophila (sympatric), discrete compatibility index, Nei's *D*, *N*=35							
Snowball	13.598	0.147	0.358	0.042	0.044	−13.021	
Linear	14.381	0.666	0.248	0.251	0.368	−5.776	
Slowdown	14.381	0.666	0.248	0.251	0.368	−5.872	0.273
Drosophila (allopatric), discrete compatibility index, Nei's *D*, *N*=34							
Snowball	23.439	0.001	0.234	2×10^{-5}	0.001	−4.527	
Linear	23.242	0.098	0.444	0.018	0.039	−3.150	
Slowdown	23.242	0.098	0.444	0.018	0.039	−3.19	0.06
Drosophila (combined), discrete compatibility index, Nei's *D*, *N*=69							
Snowball	40.303	0.001	0.272	2×10^{-4}	4.0×10^{-4}	−10.34	
Linear	42.458	0.274	0.158	0.078	0.179	−5.635	
Slowdown	39.273	0.0159	0.767	0.022	0.915	−3.960	−0.900
Lepidoptera, discrete viability index, Nei's *D*, *N*=69							
Snowball	11.133	10^{-9}	10^{-14}	0.196	4×10^{-7}	−4.26	
Linear	24.051	4×10^{-5}	2×10^{-6}	0.005	0.091	−0.861	
Slowdown	7.092	0.203	0.194	0.350	0.195	−0.213	−1.186
Lepidoptera, discrete overall compatibility index, Nei's *D*, *N*=51							
Snowball	32.656	3×10^{-5}	7×10^{-7}	0.151	10^{-4}	−13.310	
Linear	31.847	0.0530	0.068	0.147	0.012	−5.047	
Slowdown	30.197	0.221	0.080	0.087	0.271	−8.131	2.501
Frogs, compatibility (EH)×(MET), Nei's *D*, *N*=89							
Snowball	495.13	10^{-4} [1]	–	–	0.049[2]	−1.348	
Linear	341.68	0.041[1]	–	–	0.208[2]	−2.401	
Slowdown	284.441	0.072[1]	–	–	0.311[2]	−38.316	51.514
Frogs, discrete compatibility index (1−*IPO*2), Nei's *D*, *N*=139							
Snowball	257.840	0.0002	3×10^{-5}	0.001	2×10^{-6}	−38.260	
Linear	253.201	0.0037	0.0066	0.0009	0.0002	−10.710	
Slowdown	253.201	0.0369	8×10^{-14}	0.0014	2×10^{-16}	−47.901	40.142
Birds, discrete compatibility index, $\Delta T_{50}H$, *N*=132							
Snowball	76.335	0.007	0.002	0.102	0.148	−0.0463	
Linear	61.234	0.0136	0.002	0.004	0.428	−0.234	
Slowdown	61.234	0.0136	0.002	0.004	0.428	−0.234	10^{-6}
Birds, discrete compatibility index, γ-*HKY*-corrected % *cytB* divergence, *N*=108							
Snowball	40.845	4×10^{-10}	2×10^{-13}	10^{-6}	0.003	−3.141	
Linear	37.287	10^{-7}	6×10^{-9}	10^{-6}	0.382	−0.568	
Slowdown	37.287	10^{-7}	6×10^{-9}	10^{-6}	0.382	−0.568	10^{-6}

Notes:

[1]**Wilcoxon–Mann–Whitney test is used instead of a Kruskall–Wallis as there are only two categories of residuals.**

[2]**Runs test is used instead of multiple comparison test as there are only two categories of residuals.**

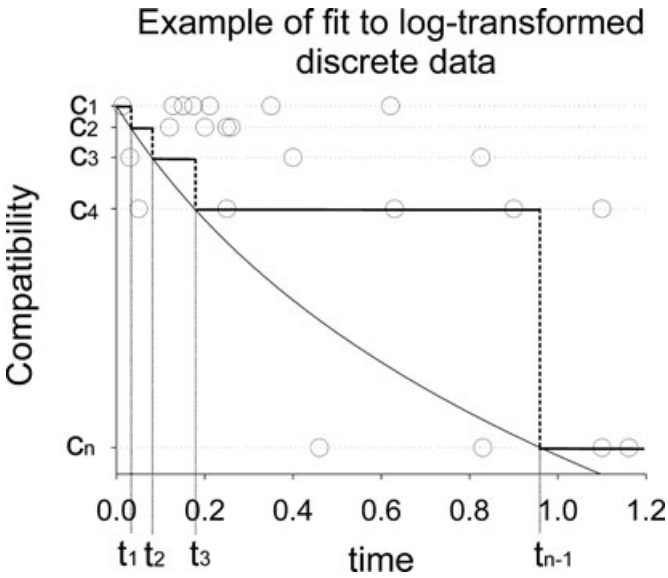


Figure 2. Method of fitting discrete data by least squares. The graph shows the method used to fit discrete compatibility data (e.g., *Drosophila* postzygotic isolation) via least squares on a log scale. Data are shown as open circles. The curve represents the original model to be evaluated. The thick lines represent the modified expectation actually used in fitting the data. The lowest expectation to be fitted is c_n, and all the $C = 0$ discrete compatibility values of the data have also been shifted to this value.

data). The theory assumes infinite sample sizes, but $C = 0$ is often realized in real data, which is finite. This is a feature of the data rather than necessarily an incorrect feature of the model (see discussion of mule sterility above). Yet these datapoints nonetheless provide potentially important information, which should not be lost. As a result, we replaced both expected $c_n = 0$ and observed $C = 0$ values to 0.001 when performing fits on log-transformed compatibility. We did this on the grounds that the data collected rarely involved sample sizes of more than 1000, so that the errors in the data are of order $>1/1000$. Shifts to different values ($C =$ 0.01 and 0.0001) were also tested; this changed the values of sums of squares, but did not strongly affect relative values of sums of squares obtained with different models.

RESIDUALS ANALYSIS FOR A QUALITATIVE CHECK OF THE GOODNESS OF FIT

Goodness of fit to the different models was investigated further by analyzing the sequence of residuals obtained with the best fit that each model can produce. We performed nonparametric analyses to test whether the distribution of residuals was equable. Minimizing the sum of squares does not ensure that a model provides good predictions throughout the range of the predictive variable, i.e., at any genetic distance, if the residuals are highly skewed. Largely positive or largely negative residuals would correspond to under- and overestimations of the model, respectively.

In the prezygotic isolation datasets, compatibility is a continuous variable, so nonzero residuals were always obtained. We performed Wilcoxon–Mann–Whitney tests to compare mean ranks of positive and negative residuals along the genetic distance axis. This tested whether the model underestimated compatibility at low genetic distances and overestimated at high genetic distances, or vice-versa. When fitting discrete postzygotic isolation measures, expected and observed values can be equal, and residuals can take the value of 0. We therefore performed Kruskall–Wallis nonparametric analyses of variances on these data. We also performed Wilcoxon–Mann–Whitney pairwise tests for differences between mean ranks of 0 versus positive (0,+), 0 versus negative (0,−), and positive versus negative (+,−) residuals.

A second nonparametric analysis was performed on all continuous data to test whether positive or negative residuals were autocorrelated. (A certain amount of autocorrelation is expected in fitting discrete compatibility data, so we performed autocorrelation tests on only on continuous compatibility data). We used a Wald–Wolfowitz runs test (where a run consists of a run of positive or negative residuals, $+ + + +$ or $- - - -$) to establish whether the model under- or overestimates compatibility in any particular range of genetic distances. It is worthwhile doing this kind of test as it is possible to imagine conditions in which the sequence of residuals is nonrandom even though mean ranks do not differ, for example if residuals are negative early and late, but positive at intermediate genetic distances, indicating a poor fit. If the smallest number of positive or negative residuals was < 10, the probability was calculated exactly; otherwise a Gaussian approximation was applied, with mean number of runs $1 + 2m(N - m)/N$ and variance $(2m(N - m)(2m(N - m)-N))/N^2(N-1)$, where m is the total number of positive residuals, and N is the total number of residuals.

Results

TESTS INVOLVING SOME DEGREE OF "PREZYGOTIC" ISOLATION

Results of the fits to each of the three models are shown in Tables 2 and 4 for untransformed and log-transformed fits, respectively. For 12 of the 14 datasets, the lowest sum of squares was obtained with the slowdown model. (Five of these support small negative values of a, and therefore fit best with a linear model tending slightly toward a convex snowball function). For nine of 14 the highest sum of squares was given by the snowball model, suggesting generally poor fits. The ratio between lowest and highest sums of squares varies greatly. Graphs of the data and fits are shown in Figure 3.

Bacillus

Because no "zygotes" are formed during in bacterial transformation, the data can be viewed as analogous to a mixture of

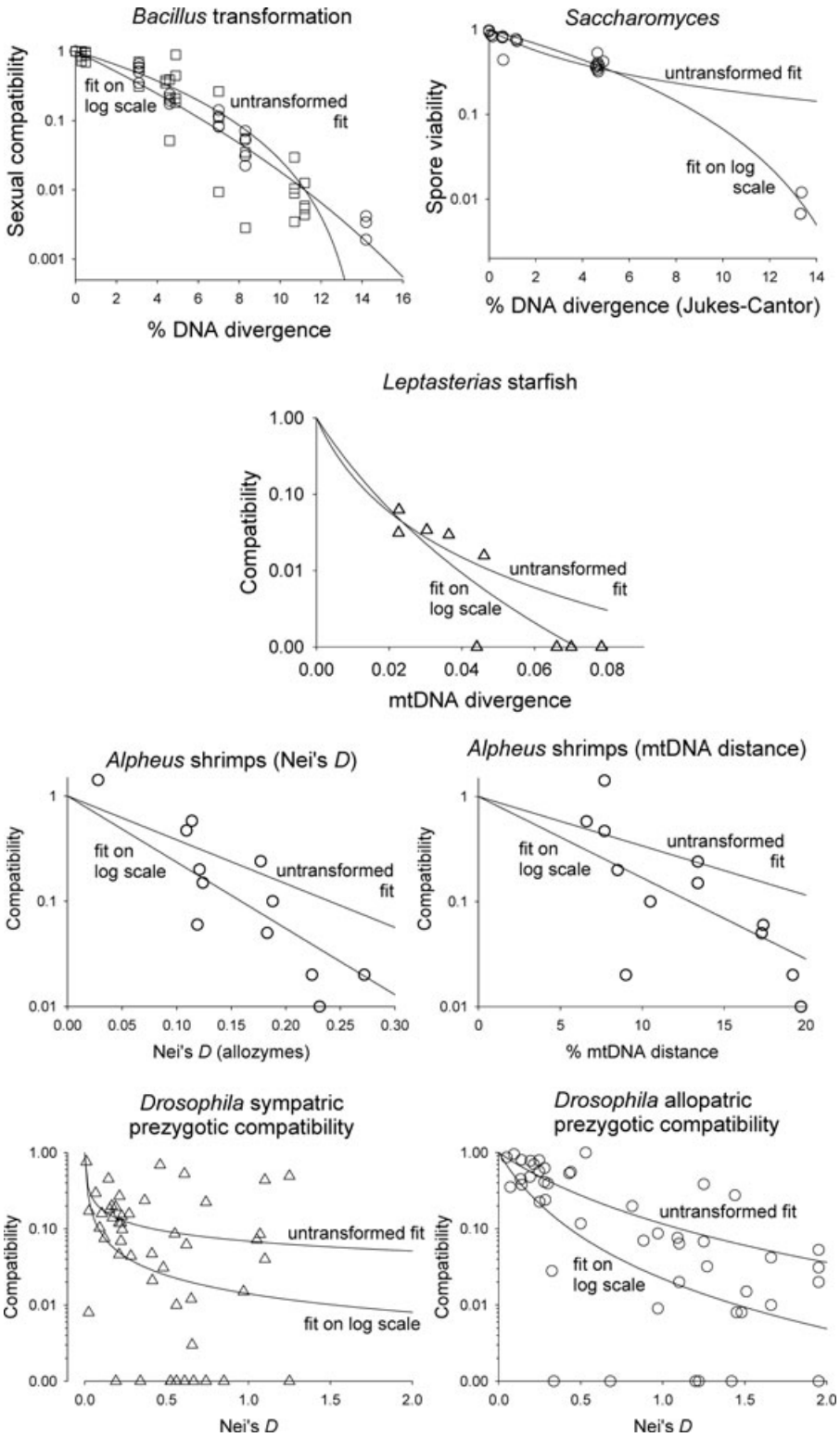


Figure 3. Fits of prezygotic and combined reproductive compatibility. In each case, compatibility is shown as a function of genetic distance (as a surrogate of time). *Bacillus*: symbols represent dilution experiments using laboratory strains (circles), and nonrestricting strains collected form the wild (squares). Otherwise, circles represent allopatric taxa or strains; triangles represent sympatric taxa or strains.

pre- and postzygotic isolation. The higher sums of squares and low *P*-values obtained using the snowball model reveal a poor fit. This is mostly due to an excess of negative residuals at low genetic distances. The observed initial decrease in reproductive compatibility is faster than expected under the snowball model. As expected, the slowdown and the linear model then do a better job as indicated by the lower sums of squares and the high *P*-values for the two nonparametric tests, especially using the logarithmic fits. The slowdown model fits the data marginally better as judged by sums of squares than the linear model, with both log-transformed and untransformed data, but only with a negative value of a (i.e., the data show a slight tendency toward snowball rather than true slowdown). This negative slowdown fit is also marginally better as judged by nonparametric tests and is shown in Figure 3, although similarities of sums of squares obtained for slowdown and linear models and the relatively weak values of parameter a are as expected if substitution rates are not very different from constant. Because we have no a priori reason to expect a snowball-like fit to be less than second order (as here), we may conclude that the null or linear model is not rejected.

Saccharomyces

Because spore viability measures the fertility of hybrids, these data might mostly be equated with postzygotic compatibility. The snowball model does not fit the data well, and fits more poorly than a slowdown model in both linear and log compatibility scales. A slowdown seems to fit the data better than the linear model on a log compatibility scale ($F = (3.449/18)/(0.953/17) = 3.42$, df = 18,17, $P = 0.01$), but only when it produces a snowball-like convex function with a negative value of a (as shown in Fig. 3); furthermore, a true slowdown ($a > 0$) fits the untransformed data best (Fig. 3), but only with very limited support in comparison to the simple linear model ($F = (0.274/18)/(0.256/17) = 1.01$, df = 18,17, $P = 0.49$). The data clearly fall into three tight genetic distance clumps, due to the relatively few lineages crossed in this group of *Saccharomyces*. Thus the data are particularly subject to phylogenetic pseudoreplication, which probably explains the poor and variable fits. It is perhaps safest to say that we cannot reject the linear model.

Starfish

As explained in Appendix B, the data are affected both by pre- and postzygotic compatibility, though prezygotic is likely uppermost. High sums of squares in both transformed and untransformed fits show that early decreases of compatibility with genetic distance are faster than expected under snowball or linear models. The slowdown model fits the rapid decrease in reproductive isolation better, whether fitted on logarithmic or linear scales (shown in Fig. 3), and low sums of squares and high *P*-values of nonparametric tests, suggest a reasonable fit. However, the data are few and the sums of squares are not significantly improved by the slowdown compared with the linear model ($F = (0.00133/8)/(0.00081/7) = 1.44$, df = 8,7, $P = 0.32$ for the untransformed compatibility fit, and $F = (9.115/8)/(5.936/7) = 1.34$, df = 8,7, $P = 0.36$ for the log transformed data), even ignoring phylogenetic correlations. The slowdown model does, however, fit the data better than the snowball model ($F = 2.54$, $P = 0.12$ for the log-transformed fit; $F = 6.02$, $P = 0.01$ for the untransformed fit).

Alpheus shrimp data

A useful feature of these data is that, because the experiments employed geminate sister species across the Isthmus of Panama, there is good phylogenetic independence, although with $N = 11$, the data are hardly extensive. The sums of squares are quite similar on both linear and log scales for all models fitted no matter what measure of genetic distance is used. Furthermore, the sequences of residuals obtained with all models are similar and the corresponding distribution of residuals is approximately random (high *P*-values throughout), no matter which measure of genetic divergence is used. To obtain such a good fit with all three models may seem surprising. This apparent paradox is explained by the low sample size, and the negative values of parameter a in the slowdown model, which predicts an acceleration rather than deceleration of substitution rates, tending toward a modest snowball-like convexity. This is consistent with the finding that observed compatibility decreases slowly at low genetic distances and then accelerates, although support for the snowball model is very weak in such a small dataset. On the whole, the null model of the simple exponential compatibility failure (fitted in Fig. 3) is not rejected.

Drosophila—allopatric species

The snowball model does not fit the data at all well, as shown by the high sums of squares and low *P*-values for the residual tests, on either log or linear compatibility scales. This is due to a more rapid decrease of compatibility with genetic distance than expected under the snowball model. Thus, the linear model fits these *Drosophila* data better and, as for the starfish, the slowdown model (shown in Fig. 3) performs marginally better with a lower sum of squares and generally larger *P*-values. However, the slowdown and linear models are not clearly distinguished, with ratios of sums of squares (*F*-ratios) in the range of only 1.01–1.06.

Drosophila—sympatric species

Here, the slowdown model (Fig. 3) provides a good fit to this prezygotic data as indicated by the low sum of squares and the high nonparametric test *P*-values obtained on both scales. Both the snowball and the linear model do a worse job with larger sums of squares with especially low *P*-values in nonparametric tests for log-transformed fits of the snowball model. Such a difference in

goodness of fit obtained with the slowdown model and the two other models is consistent with the highly positive estimated values of parameter a. Indeed, such high values of a mean a much more rapid initial decrease in compatibility with genetic distance than under either snowball or linear models. There is reasonable, although not strong evidence for slowdown versus linear models when the fit is performed on a log compatibility scale ($F = 1.62$, $P = 0.06$) and similar although weaker evidence ($F = 1.39$, $P = 0.14$) with untransformed compatibility. Once again, however, the snowball model is strongly rejected in favor of the slowdown model, but only for the log compatibility fit ($F = 2.59$, $P = 0.001$ for the logarithmic fit; $F = 1.34$, $P = 0.17$ for the linear fit).

TESTS INVOLVING ONLY POSTZYGOTIC ISOLATION

Results of the fits of the postzygotic datasets to the three models for both linear compatibility and log compatibility transformations are summarized in Tables 3 and 5, and shown graphically in Figure 4.

Drosophila—sympatric and allopatric datasets combined

Although we provide separate fits for allopatric and sympatric datasets for comparison with the prezygotic data (Tables 3 and 5), there is no evidence for differences in accumulation of postzygotic isolation (Coyne and Orr 1997; Mendelson et al. 2004). Whether we analyze the datasets together or separately, they remain obstinately insufficient for distinguishing between models, in part perhaps because of the discrete nature of the data. Therefore, we discuss here only the combined sympatric and allopatric postzygotic data. The snowball model gives a poor fit, as evidenced by the low P-values and higher sum of squares in the untransformed compatibility fit. The linear fit shows higher P-values in nonparametric residual tests, and is therefore preferred for the untransformed data, even though sums of squares are not significantly smaller ($F = 1.29$, $P = 0.15$. For the log-transformed fit, the slightly negative value of a in the preferred slowdown model indicates a weak snowball tendency ($F = 1.06$, $P = 0.41$); however, this is not true when allopatric and sympatric data are analyzed separately (Table 5). Overall, the linear model (Fig. 4) is most compatible with the data, based chiefly on higher P-values for the residual tests, on whatever scale of fitting is used. This result differs somewhat from Orr's (1995) conclusion that there was weak evidence for a snowball effect in the *Drosophila* data. However, Orr's conclusion was tentative, and was made only by comparing reproductive isolation in one direction of cross with that in both directions, as opposed our own curve-fitting analysis of the whole data, which also takes into account the discrete values of reproductive isolation.

Lepidoptera

In this dataset, the viability data on their own appear to fit a snowball model better than linear ($F = 1.30$, $P = 0.14$ for untransformed compatibility fit, $F = 2.16$, $P = 0.001$ on a log scale) and also better than a slowdown model on untransformed data ($F = 1.32$, $P = 0.13$). The most strongly supported model for log-transformed compatibility is a reverse slowdown model (i.e., snowball-like, with negative a), even against the next-best snowball model ($F = 1.55$, $P = 0.04$), and also against the linear model ($F = 1.55$, $P < 0.0001$). This snowball-like model on a log compatibility scale (Fig. 4) was also much the best fit in terms of nonparametric residuals tests, but it only performs this fit so well by becoming undefined with genetic distances > 0.84. In contrast, the snowball was a better fit according to nonparametric tests with untransformed compatibility data as compared with linear or slowdown models, even though only weakly supported via sums of squares and the F ratio. We tentatively suggest that there is some evidence for a snowball model from this data. On the other hand, when total postzygotic incompatibility (which includes hybrid fertility as well as viability) is assessed, the situation reverses, the snowball model develops low P-values in nonparametric tests, whereas linear (Fig. 4) and slowdown models become more supported, although only weakly ($F \approx 1$). In conclusion, the data from Lepidoptera are mixed: it is possible that viability evolves according to the snowball model, whereas viability + fertility considered together do not clearly support any model. The major reason for these mixed results is the large amount of scatter in the data. Whereas some pairs of species can become completely incompatible with Nei's D as low as 0.1, others remain viable and fertile in at least one direction of cross with Nei's $D > 0.8$.

Frogs: egg hatch and metamorphosis compatibility

The snowball model appears to provide a poor fit to the data compared to the linear whether viewed on linear or log compatibility scales ($F = 1.30$, $P = 0.11$; $F = 1.45$, $P = 0.04$, respectively), and low P-values of non-parametric tests on the log scale support this rejection of the snowball model. The slowdown model (Fig. 4) gives a still better fit than linear, although it is not much better ($F = 1.19$, $P = 0.21$; $F = 1.25$, $P = 0.15$, respectively).

Frogs: discrete compatibility index (1-IPO2)

In these data, only three discrete values are possible, 0.0, 0.5, and 1.0. Perhaps unsurprisingly there is therefore much scatter around the best fit lines, very low P-values for the nonparametric tests, and little way to distinguish between models in terms of sums of squares. Overall, a linear model is not rejected by these data (Fig. 4), although untransformed data appear marginally to support a slowdown model, and log-transformed data appear marginally to support a linear or slowdown model (the latter two being indistinguishable due to the discrete values of the data).

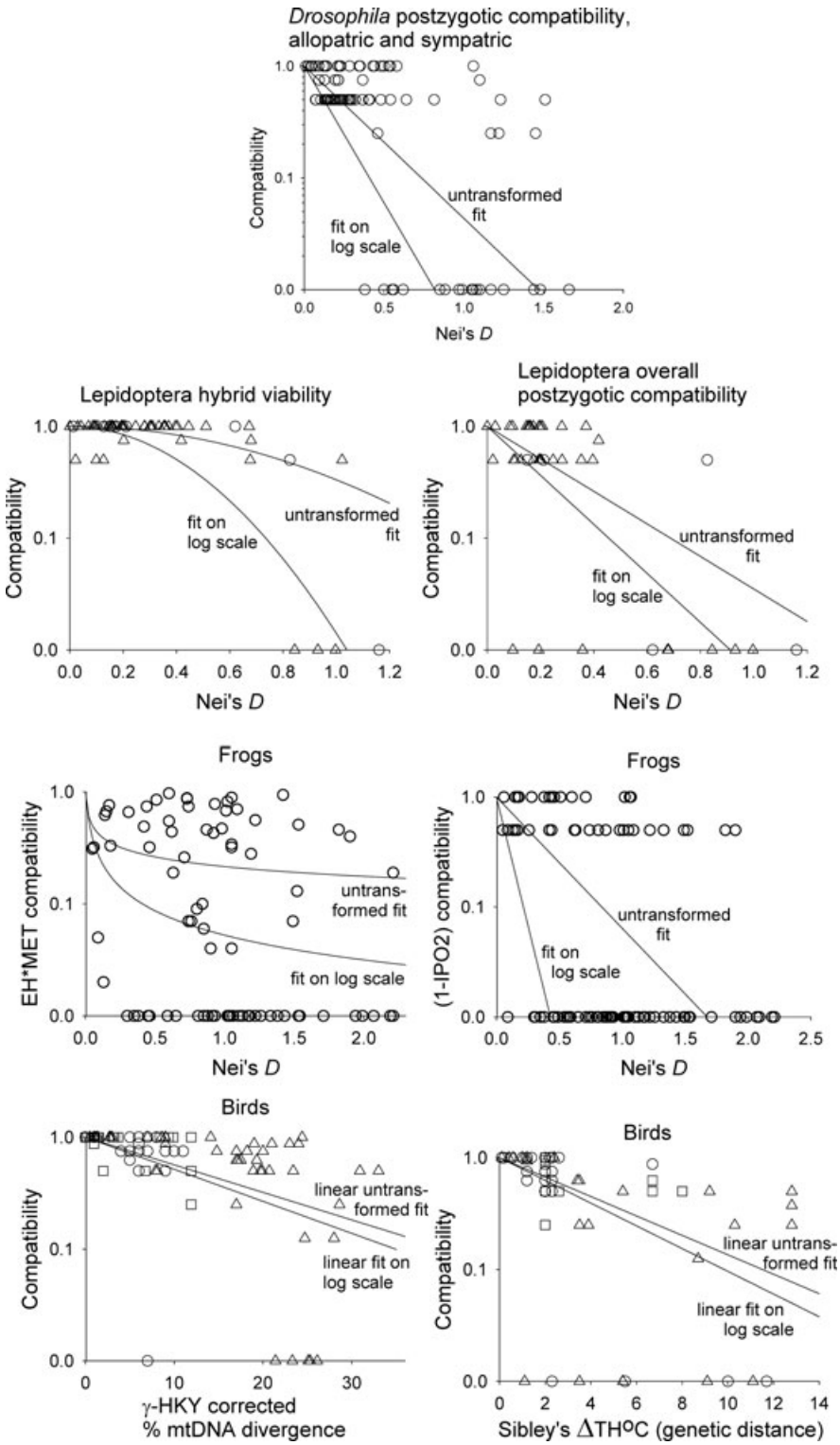


Figure 4. Fits of postzygotic compatibility. Symbols for Lepidoptera represent allopatric (circles) and sympatric (triangles) taxa. For birds, squares represent duck data, circles represent passerine data, and triangles represent other nonpasserine, nonduck data.

Birds

The best fit is either linear or does not differ significantly from the linear model in every test (Fig. 4). It should also be noted that none of the models fit very well, as judged by the residuals tests. The reason for this is clear when looking at plots of the data, which reveal enormous scatter. Although some species become completely incompatible via sterility and/or hybrid inviability by around 2% *cytB* or 1°C $\Delta T_{50}H$ distances, other pairs of

species may remain almost fully viable and fertile (>75%) until distances of 24% divergence of *cytB* or 7°C $\Delta T_{50}H$. It is tantalizing that some of this scatter might be due to differences in the rate of evolution of incompatibility between different bird groups. The data for passerines and ducks that can be crossed all refer to low genetic distances, probably in part because both represent mostly recent groups that are relatively homogeneous, but also possibly because these groups evolve strong incompatibilities early, leading to an impossibility of making crosses across large genetic divergences. The crosses among nonduck, nonpasserine birds, on the other hand, are mostly between relatively genetically distant species (>10% *cytB*), although the few crosses among closer species in this category do not give convincing evidence for inhomogeneity with passerines and ducks. In conclusion, the bird data are somewhat unsatisfactory, in that no model is supported strongly, nor is any model rejected strongly, and this may in part be due to inhomogeneity of rates of incompatibility evolution between different groups of birds. In general, however, there is no clear evidence for deviation from the linear model, and the overriding impression from the bird data is of a great deal of scatter.

Discussion

In this article, we test the time course of compatibility evolution using three simple phenomenological models, justifying these in part via constant and variable substitution rates, and considering accumulation of either nonepistatic or two-locus epistatic Dobzhansky–Muller incompatibilities, and multiplicative combination of different incompatibilities. Under these models, together with multiplicative fitness combinations, the decrease in log overall compatibility with time is expected to correspond to a linear, concave (slowdown) or convex (snowball) function of genetic distance. In the two latter cases, the initial decrease in compatibility between species at low genetic distances is expected to be respectively faster or slower than linear, compared to later decreases.

There are many potential statistical problems with fitting such data (Appendix C). Our exploratory approach (Mosteller and Tukey 1977) is perhaps not very powerful for confirmatory data analysis, and there are also difficult problems in this kind of data due to phylogenetic correlations of unknown extent (Bolnick and Near 2005). The data themselves are often fragmentary, and as different methodologies were used in each study of our meta-analysis one should be cautious about reaching firm conclusions. There is also the problem of using genetic distance as a proxy for time since speciation (Bolnick and Near 2005). Nonetheless, it is worth exploring for any potential patterns in such data, which can then be investigated further.

EVIDENCE FOR REINFORCEMENT

One of our findings is that "prezygotic" datasets (in which we include the *Bacillus*, *Saccharomyces*, and *Leptasterias* data) often support the slowdown model, although normally only weakly, and that the snowball model is generally rejected for such data, as expected (because the snowball model was developed only to explain inviability and sterility, as in Haldane's rule—Orr 1995). Two prezygotic datasets stand out in their support for slowdown: the sympatric *Drosophila*, and *Leptasterias* (starfish). The starfish results are intriguing in that the data are in the form of numbers of hybrids compared with numbers of much commoner parentals in a sample from nature. The starfish thus provide some of the few quantitative data on successful interspecific hybridization from the wild (for a survey of other data, see Mallet 2005). Assuming that F_1 hybrid zygotes develop as readily as pure species once fertilization has taken place, this is one of the few cases in which complete prezygotic compatibility between species has been measured in the wild, with all the effects due to habitat separation, timing of spawning, and fertilization probability integrated into the measure. Some laboratory studies of these starfish have shown that hybrids are readily produced given fertilization (Foltz 1997), but unfortunately we cannot entirely rule out the possibility that some of the rarity of the field-sampled hybrids is due to developmental or ecological problems in the hybrids, and therefore the data will probably contain postzygotic as well as prezygotic information. In addition, because all the species are closely related, and many of the species were used more than once, there will be problems of phylogenetic pseudoreplication in the data, which are anyway based on small numbers of species. Nonetheless, the support for slowdown, exactly as expected if reinforcement were to occur to limit hybridization rather soon after speciation, presumably via changes in prezygotic compatibility or the time and place of gamete release, remains of great interest. The results, put simply, show that there are too few hybrids between species of low genetic distance to be explained by a simple exponential failure of compatibility. These data highlight a hitherto underused method to test for reinforcement in other species for which they and their hybrids might be sampled in nature.

The second example for which slowdown provides the best fit is in laboratory prezygotic data for sympatric *Drosophila* species. Here, there is more scatter than in the starfish, but also considerably more data. Although evidence for slowdown seems fairly clear on the log compatibility scale, it should be remembered that our tests do not allow for phylogenetic correlations, which would reduce the degrees of freedom by around 20%–50% (Coyne and Orr 1997). Nonetheless, the data do appear to conform best to the slowdown model, due to very rapid initial, and then decelerating acquisition of prezygotic incompatibility. Previously, the more rapid average acquisition of prezygotic isolation of sympatric than of allopatric *Drosophila* species suggested reinforcement as

a likely cause (Coyne and Orr 1989, 1997, 2004). Here, we identify a new and hitherto unrecognized trait in the same sympatric data, which is also indicative of reinforcement, that of concave curvature of the fit or deceleration on a log scale. Together, these findings on sympatric *Drosophila* implicate the evolution of assortative mating as a result of selection in incompletely reproductive isolated populations in sympatry.

THE CASE OF THE MISSING SNOWBALL

The theory of negatively epistatic "Dobzhansky–Muller" incompatibilities (Orr 1995; Orr and Turelli 2001) is now well-founded and widely accepted as a major cause of incompatibility, particularly in Haldane's rule (Turelli et al. 2001; Coyne and Orr 2004; Mallet 2006; Johnson 2006). A corollary is that one expects an accelerating rate of incompatibility accumulation (Orr 1995), the so-called snowball effect. If exactly two epistatic loci are involved in each incompatibility, incompatibilities should accumulate as a quadratic function, although there is no reason why three, four, or more loci might not be involved, in which case the curvature of incompatibility accumulation would have correspondingly higher power. Although we here model only a quadratic snowball, more convex curvatures should still be fit better by a quadratic function rather than by the nearest, linear alternative we test.

Although a number of attempts have been made to fit reproductive isolation versus genetic distance (Edmands 2002; Fitzpatrick 2002; Mendelson et al. 2004; Bolnick and Near 2005), very little evidence has been seen for quadratic or higher order acceleration of reproductive isolation predicted by the snowball model. This problem has been dubbed "the missing snowball" (Johnson 2006). Even the authors of the original *Drosophila* comparative data paper fitted a log-linear (i.e., simple exponential failure) model to their data (Coyne and Orr 1997), rather than the snowballing model originally motivated by these same data. On the other hand, methodologies hitherto used in fitting such data were simple linear regression fits, and often did not even constrain reproductive isolation to be zero in the absence of genetic divergence. The effect of multiple incompatibilities on fitness has not been modeled previously. Others have discussed the problem of mapping incompatibilities onto fitness, but treat speciation or reproductive isolation as a threshold trait that requires a certain number of additive incompatibilities. Their theory of speciation is that when the numbers of incompatibilities reach this threshold, reproductive isolation and speciation is complete (Orr and Turelli 2001; Turelli and Moyle 2007). In retrospect, it seems odd that a more general population genetic multiplicative fitness approach (i.e., $\Pi(1 - s_i)$, where s_i represents the selection coefficient due to the ith incompatibility) was not used (Walsh 1982; Orr 1995; Gavrilets 2004). Turelli and Moyle (2007), noting the missing snowball in empirical studies, suggested that the problem might be due to incompatibilities having diminishing deleterious effects as the numbers of incompatibilities approached the threshold, so giving a more linear rate of accumulation of reproductive isolation overall. However, we can think of no good a priori reason why the effects of incompatibilities should decrease in this way, and our multiplicative fitness approach is the simplest to give a fitness curvature similar to their diminishing effect model while retaining the full proportional effects of every incompatibility. Turelli and Moyle (2007) further argue that their additive fitness scale is in any case appropriate because it approximates multiplicative fitness when s_i values are low: while this is true, this additive approximation breaks down when very many such incompatibilities are considered together, as they must be when describing the entire spectrum of incompatibility evolution, as here.

It is of interest that, even after multiplicative fitness incorporation, the snowball model still fails to provide the best fit for postzygotic (or prezygotic) datasets. The only possible exceptions are the Lepidoptera viability-only data. The snowball predictions not met are (1) that compatibility should decline slowly at first, but then faster later during incompatibility evolution, and (2) that the variance is very low at the beginning of the process compared with other models. Both of these characteristics can be clearly seen in the simulated data of Figure 1B. Although the snowball model is based on epistatic interactions at only two loci, the same features will be even more extreme for complex epistasis involving three or more loci. The slow start and low initial variance both result from the fact that very few incompatibilities arise early during divergence. Each incompatibility must result from two or more "hits," or mutations at two or more separate interacting substitutions. This low initial variance effect forms a strong contrast with the linear and slowdown models that require only a single "hit" to form an incompatibility and are expected to accumulate initial variance much more rapidly, due to vagaries of the mutation process (see Fig. 1). It is important to note that, while it is in a sense true that "reproductive isolation must ... increase faster than linearly with time" under the snowball model (Orr 1995), "faster" here refers to the acceleration, rather than to the average rate of acquisition of reproductive isolation. The acquisition of a given amount of reproductive isolation depends on details such as relative substitution rates of epistatic and nonepistatic effects and their relative strengths, and is likely to be slower at first for epistatic reproductive isolation than for nonepistatic effects.

The data provide a poor fit to snowball theory primarily because some species evolve high levels of incompatibility extremely rapidly whereas other pairs remain compatible for a long time. Our other major finding is that there is often extremely high variance in the data, so that the second prediction of very low variance early in divergence is not met either (although we did not incorporate variance in our final fitting procedure). Bearing

these two deviations from snowball predictions in mind, a number of explanations are possible.

(1) The multiplicative fitness assumption for incompatibility combination is overly restrictive

Other models of hybrid unfitness accumulation are possible, and might be tuned to fit the data (e.g., Turelli and Moyle 2007). We argue that multiplicative fitness is the simplest and most reasonable starting point for such models, especially where there is a lack of evidence against it, and it is also a standard in population genetics. Furthermore, as individual Dobzhansky–Muller incompatibilities require at least two changes, we can make the strong prediction of few early incompatibilities, leading to very low variance during this early phase (Fig. 1B), whatever the deviation from multiplicative fitness. Tinkering with the fitness function of early and late incompatibilities will not alter this lack of correspondence with the data.

Reviewers of an earlier version of this article have claimed that it is impossible to test predictions of the snowball model in the way we have done on the grounds that "nothing is known about the relationship between the number of incompatibilities and the decline in hybrid fitness" (see also Note 1, above). A simple way to model more general fitness accumulation among multiple incompatibilities might be to introduce a term for epistasis, ε (note, this is different from the usual meaning of "epistasis," which, as in the positive epistasis in Dobzhansky–Muller incompatibilities, is usually reserved for fitness effects of multiple genes). The fitness of an individual affected by two incompatibilities, i and j, would then be: $(1 - s_i)(1 - s_j) - \varepsilon_{ij}$. Most of the discussion about fitness combination of deleterious mutations has been about whether positive epistasis is or is not observed (Charlesworth et al. 2004). If ε is positive, the effect of more mutations will be even more extreme than multiplicative, and our convex snowball curves should fit more and more data, even of non-Dobzhansky–Muller incompatibilities, rather than hardly any, as we find in this article. To force Dobzhansky–Muller type incompatibilities into more approximately nonsnowball log-linear fitness declines, as we find in the data, ε would have to be negative. (Note that the "absolute" fitness effects of further incompatibilities do in fact decline as more incompatibilities are fixed in the multiplicative model—it is only the "proportional" fitness effects that remain constant). We can think of no a priori reason why epistasis among incompatibilities should be negative, especially as it reverses the normal Dobzhansky–Muller epistasis of substitutions within incompatibilities. Negative epistasis would imply that escape from death via a genetic incompatibility early in life predisposes an individual to successful escape from a different genetic incompatibility acting later. If incompatibilities are independent, this seems unlikely; assuming $\varepsilon \geq 0$ therefore seems reasonable (Walsh 1982; Orr 1995).

(2) Many different snowball-like processes will give more complex curves than a simple snowball process

For example, suppose that in *Drosophila* each of four sequential processes completes in hybrids before the next starts: (1) Haldane's rule male sterility; (2) Haldane's rule male inviability; (3) bisexual sterility, (4) bisexual inviability. Then although each process may proceed via a perfect quadratic snowball, when placed end to end the overall function could appear more or less linear. With a little stochastic variation, and some inevitable overlap between the four processes, the results would be difficult to distinguish from linear. This suggestion seems a very likely cause of at least some of the poor fit of the snowball process for overall hybrid fitness in *Drosophila*, Lepidoptera, frogs, and birds, as well as other organisms for which postzygotic isolation has been assessed in which many complex processes are probably involved (particularly those involving Haldane's rule and sterility as well as inviability). The stronger support for the snowball we have found for viability measured on its own in Lepidoptera, coupled with the better fits with overall hybrid fitness for a linear model weakly support this interpretation. However, multiple processes will not explain the lack of snowball fits in *Bacillus* or *Saccharomyces*, which have less predicted complexity.

(3) High variation in substitution rate or incompatibility accumulation among lineages

Given that we know very little about the processes governing the evolution of negative epistatic incompatibilities (Welch 2004), this seems a possible explanation for some of the scatter in the data. However, although somewhat variable, DNA sequence evolution with few fitness effects on its own genetic backgrounds does not tend to deviate very widely from an approximate molecular clock. Variation in average substitution rates and compatibility decline alone therefore seems unlikely to explain the large scatter in incompatibility early in divergence, as seen in the data. It is more likely due to stochastic evolution of genes with major effects (see 5).

On the other hand, it is likely that different groups of species accumulate incompatibilities at different rates, in spite of similar rates of DNA substitution. Something of this type may be occurring in the rapidly radiating passerines and ducks, compared with other bird species (see discussion of the bird data above for more details).

(4) Incompatibility mostly evolves linearly, which overcomes the signal due to snowball epistasis

This proposal seems unlikely to explain all of the data, because of the well-established nature of Dobzhansky–Muller incompatibility theory (Welch 2004), especially in explaining Haldane's rule and the evolution of other complex incompatibilities in organisms such as *Drosophila*. Nonetheless, the theory is not ruled

out for some other sorts of incompatibilities. If chromosomal evolution, for example, were to occur via occasional drift in small populations (Walsh 1982; Gavrilets 2004), or equivalently, during unusual bouts of positive selection, and local population sizes remained roughly similar over evolutionary time, chromosomal incompatibilities might accumulate roughly linearly. Local evolution of inversions by adaptation-trapping (Kirkpatrick and Barton 2007) might also result in an approximately linear accumulation of incompatibilities with time, because that theory again does not require epistasis. Although earlier drift-based theories of chromosomal evolution and speciation have been all but ruled out (Coyne et al. 1991; Coyne and Orr 2004), it does not seem improbable that chromosomal rearrangements sometimes contribute to strong hybrid sterility, given frequent observations of chromosomally based sterility in mammals (Chandley 1988; Britton-Davidian et al. 2000). In *Saccharomyces* yeasts, a strong effect of chromosomal rearrangements alone on fertility of diploid hybrids among species has been demonstrated by reverse-engineering the rearrangements while leaving epistatic effects intact (Delneri et al. 2003).

Chromosomal evolution is not the only linear evolutionary process that leads to reproductive isolation. In *Saccharomyces* and *Bacillus*, a further process seems to be at work, due to a direct negative effect of divergence on recombination; recombination is important for successful meiosis (yeasts) or transformation (*Bacillus*). This process readily explains the slow, approximately linear accumulation of incompatibility in both *Saccharomyces* and *Bacillus* datasets. In principle, sequence divergence should also contribute directly to incompatibilities in the higher eukaryotes. However, to reduce compatibility by 99% requires overall sequence divergence of $>10\%$ in both microbial datasets, so it is possible that other processes, such as the snowball, are more important when multicellular eukaryotes evolve incompatibilities at lower genetic divergences. Nonetheless, it should not, perhaps, be ruled out that chromosomal and other "single-hit" incompatibilities might contribute to reproductive isolation at high enough rates to be significant compared to those causing Dobzhansky–Muller incompatibilities.

(5) Incompatibilities typically have major and highly variable effects (high and variable s_i)

In our formulation we have assumed low and constant s_i, so poor fit to the snowball is readily explained by deviation in the data from this assumption. Because Dobzhansky–Muller incompatibilities are caused by epistatic interactions previously untested by natural selection before they are expressed in hybrids, there is no reason why their effects should not be major on the hybrid background. Genes of major effect would cause so much scatter in the data that an underlying snowball curvature might be indistinguishable. A growing number of genes are now known that are highly deleterious to hybrids in *Drosophila* (reviewed by Wu et al. 1996; Orr et al. 2004; Mallet 2006), and also some other species such as fish (Wittbrodt et al. 1989) or Lepidoptera (Naisbit et al. 2002). Recently, the striking pattern of asymmetric incompatibilities, as commonly observed in reciprocal crosses (e.g., *A* female $\times$ *B* male produces fertile hybrids, whereas *B* female $\times$ *A* male produces sterile hybrids), has been investigated theoretically: one of the most likely explanations is the stochastic accumulation of very variable and large fitness effects via epistatic genes (Turelli and Moyle 2007). We therefore regard stochastic accumulation of major-effect substitutions as a likely explanation for much if not all of the scatter, and the poor fit with most models, including the snowball model, in higher eukaryotes.

Conclusions

The species boundary will be crisp and "real" if compatibility between populations declines precipitously, and at an accelerating rate during speciation. Motivated by a desire to test this prediction of Orr's (Orr 1995) snowball model of accelerating overall incompatibility accumulation, we propose two alternatives suitable for fitting to comparative data. The first of these, the "exponential failure law" (which we call the linear model here), is the usual model in mechanical or light-bulb failure and simple survival curves, in which incompatibilities accumulate linearly with time. The hypothesis that the evolution of premating isolation may be driven by gene flow (i.e., reinforcement), with diminishing effects as isolation increases, motivates our second alternative, the slowdown model. All three models were mapped onto fitness via a multiplicative fitness scheme.

We then test these models against empirical comparative studies of reproductive isolation accumulation. We find some evidence for slowdown, i.e., decelerating incompatibility accumulation, as expected under reinforcement, in the data for sympatric *Leptasterias* starfish and sympatric *Drosophila* prezygotic isolation. Existing data are not extensive enough to prove this hitherto untested prediction of reinforcement theory beyond a shadow of doubt. However, it is encouraging to find a slowdown pattern at all, roughly where we expect it in sympatric species pairs, and our finding suggests that further investigation into slowdown effects may be worthwhile.

Under the two-locus snowball model, first, mean compatibility is expected to remain high early in divergence, second, a very low initial variance of compatibility is expected, and third, incompatibilities are expected to accumulate rapidly once the process has started. All three are due to quadratic incompatibility accumulation, and will be more extreme for greater complexity of epistasis (three or more genes). The data, in contrast, often suggest rapid initial compatibility loss, and a very high degree of scatter early in divergence, as well as late. This "missing

snowball" (Johnson 2006), especially from where it is expected in postzygotic compatibility of higher eukaryotes, is probably best explained not because Dobzhansky–Muller incompatibilities do not occur, but by a combination of (1) stochasticity caused by a few genes having major and variable effects, and (2) a number of different overlapping snowball processes occurring at widely different rates, for example, Haldane's rule sterility and inviability versus bisexual sterility and viability (Turelli and Orr 1995). Although Dobzhansky–Muller incompatibilities undoubtedly occur, they do not seem to lead to accelerating loss of overall compatibility. Studies with the microorganisms *Bacillus* and *Saccharomyces*, as well as consideration of chromosome evolution, suggest that some incompatibility accumulation may also be truly linear, giving rise to genetic differences that produce simple exponential compatibility failure. It seems possible that some such processes, due to nonepistatic incompatibilities (such as selection against chromosomal heterozygotes), may also be involved to a nontrivial extent in macro-organismal incompatibilities.

Perhaps the major pattern indicated by these data from multicellular organisms is the one originally noted by Darwin, that hybrid inviability and sterility is indeed associated with speciation, but that its variability among different pairs of species implies only a loose association (Darwin 1859). In birds and centrarchid fish, sister species often remain at least partially compatible and able to exchange genes for many millions of years after speciation (Price and Bouvier 2002; Bolnick and Near 2005), and the data surveyed here show this to be rather general. Furthermore, there is great variability in the rates of accumulation of postzygotic incompatibility. In contrast, rapid evolution of prezygotic isolation, via "slowdown" evolution of assortative mating among sympatric populations, seems to provide a clearer species boundary, at least for sympatric taxa, than the slow and highly variable evolution of inviability and sterility.

ACKNOWLEDGMENTS

We are grateful to support from NERC, BBSRC, and DEFRA-Darwin Initiative and to the European Commission for a Marie Curie post-doctoral fellowship to SG (HPMF-CT-2001-01230) during the course of this work. We thank M. Turelli, D. Greig, Z. Yang, D. Presgraves, M. Noor, and a number of anonymous reviewers for discussions about earlier versions of this article.

LITERATURE CITED

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei : Centrarchidae). Evolution 59:1754–1767.

Britton-Davidian, J., J. Catalan, M. da Graça Ramalhinho, G. Ganem, J. C. Affray, R. Capela, M. Biscoito, J. B. Searle, and M. da Luz Mathias. 2000. Rapid chromosomal evolution in mice. Nature (London) 403: 158.

Butlin, R. K. 1995. Reinforcement: an idea evolving. Trends Ecol. Evol. 10:432–434.

Chandley, A. C. 1988. Fertile mules. J. R. Soc. Med. 81:2.

Charlesworth, B., H. Borthwick, C. Bartolomé, and P. Pignatelli. 2004. Estimates of the genome mutation rate for detrimental alleles in *Drosophila melanogaster*. Genetics 167:815–826.

Chen, W., and S. Jinks-Robertson. 1999. The role of the mismatch machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. Genetics 151:1299–1313.

Coyne, J. A., S. Aulard, and A. Berry. 1991. Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. Genetics 129:791–802.

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

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

———. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51:295–303.

———. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Delneri, D., I. Colson, S. Grammenoudi, I. N. Roberts, E. J. Louis, and S. G. Oliver. 2003. Engineering evolution to study speciation in yeasts. Nature (London) 422:68–72.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. Am Nat 74:312–321.

Dopman, E. B., S. M. Bogdanowicz, and R. G. Harrison. 2004. Genetic mapping of sexual isolation between E and Z strains of the European corn borer (*Ostrinia nubilialis*). Genetics 167:301–309.

Drès, M., and J. Mallet. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. Philos. Trans. R. Soc. Lond. B 357:471–492.

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

Fitzpatrick, B. M. 2002. Molecular correlates of reproductive isolation. Evolution 56:191–198.

Foltz, D. W. 1997. Hybridization frequency is negatively correlated with divergence time of mitochondrial DNA haplotypes in a sea star (*Leptasterias* spp.) species complex. Evolution 51:283–288.

Gavrilets, S. 2003. Models of speciation: what have we learned in 40 years? Evolution 57:2197–2215.

———. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.

Greig, D. 2007. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. PLoS Genet. 3:e21.

Greig, D., M. Travisano, E. J. Louis, and R. H. Borts. 2003. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. J. Evol. Biol. 16:429–437.

Higgie, M., S. Chenoweth, and M. W. Blows. 2000. Natural selection and the reinforcement of mate recognition. Science 290:519–521.

Hollocher, H., and C. I. Wu. 1996. The genetics of reproductive isolation in the *Drosophila simulans* clade: X *vs.* autosomal effects and male *vs*. female effects. Genetics 143:1243–1255.

Hunter, N., S. R. Chambers, E. J. Louis, and R. H. Borts. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. 15:1726–1733.

Jiggins, C. D., I. Emelianov, and J. Mallet. 2005. Assortative mating and speciation as pleiotropic effects of ecological adaptation: examples in moths and butterflies. Pp. 451–473 *in* M. Fellowes, ed. Insect evolutionary ecology. Royal Entomological Society, London.

Johnson, N. A. 2006. Patterns and processes of speciation: the evolution of reproductive isolating barriers. Pp. 374–386 *in* C. W. Fox and J. B.

Wolf, eds. Evolutionary genetics. concepts and case studies. Oxford Univ. Press, Oxford.
Kirkpatrick, M., and N. Barton. 2007. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.
Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection. Am. Nat. 159:S22–S35.
Knowlton, N., L. A. Weigt, L. A. Solórzano, D. K. Mills, and E. Bermingham. 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science 260:1629–1632.
Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. USA 99:14878–14883.
Liti, G., D. B. H. Barton, and E. J. Louis. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics 174:839–850.
Majewski, J., and F. M. Cohan. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. Genetics 148:13–18.
Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–237.
———. 2006. What has *Drosophila* genetics revealed about speciation? Trends Ecol. Evol. 21:386–393.
Mallet, J., W. O. McMillan, and C. D. Jiggins. 1998. Mimicry and warning color at the boundary between races and species. Pp. 390–403 *in* D. J. Howard, ed. Endless forms: species and speciation. Oxford Univ. Press, New York.
Mendelson, T. C., B. D. Inouye, and M. D. Rausher. 2004. Quantifying patterns in the evolution of reproductive isolation. Evolution 58:1424–1433.
Mosteller, F., and J. W. Tukey. 1977. Data analysis and regression. Addison-Wesley, Reading, MA.
Naisbit, R. E., C. D. Jiggins, M. Linares, and J. Mallet. 2002. Hybrid sterility, Haldane's rule, and speciation in *Heliconius cydno* and *H. melpomene*. Genetics 161:1517–1526.
Noor, M. A. F. 1995. Speciation driven by natural selection in *Drosophila*. Nature (London) 375:674–675.
Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.
———. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949.
Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55:1085–1094.
Orr, H. A., J. P. Masly, and D. C. Presgraves. 2004. Speciation genes. Curr. Opin. Genet. Dev. 14:675–679.
Presgraves, D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168–1183.
Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.
Read, A. F., and S. Nee. 1991. Is Haldane's Rule significant? Evolution 45:1707–1709.
Rieseberg, L. H., B. Sinervo, C. R. Linder, M. C. Ungerer, and D. M. Arias. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. Science 272:741–745.
Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811–1820.
Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation: theory and data. Annu. Rev. Ecol. Syst. 34:339–364.
True, J. R., B. S. Weir, and C. C. Laurie. 1996. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics 142:819–837.
Turelli, M., and L. G. Moyle. 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's Rule. Genetics 176:1059–1088.
Turelli, M., and H. A. Orr. 1995. The dominance theory of Haldane's rule. Genetics 140:389–402.
———. 2000. Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154:1663–1679.
Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16:330–343.
Walsh, J. B. 1982. Rate of accumulation of reproductive isolation by chromosome rearrangements. Am. Nat. 120:510–532.
Welch, J. J. 2004. Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. Evolution 58:1145–1156.
Wittbrodt, J., D. Adam, B. Malitschek, W. Maueler, F. Raulf, and A. Telling. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. Nature (London) 341:415–421.
Wu, C. I. 1996. Now blows the east wind. Nature (London) 380:105–106.
Wu, C. I., M. F. Palopoli, and N. A. Johnson. 1996. Haldane's rule and its legacy: why are there so many sterile males? Trends Ecol. Evol. 11:281–284.
Zawadzki, P., M. S. Roberts, and F. M. Cohan. 1995. The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* is robust. Genetics 140:917–932.

Associate Editor: M. Rausher

Appendix A

EXPECTATION AND VARIANCE OF NUMBERS OF SUBSTITUTIONS, INCOMPATIBILITIES, AND OVERALL COMPATIBILITY WITH TIME UNDER VARIOUS MODELS

Expectation and variance of the number of substitutions K_T

Straightforward calculations show that, considering equation (2), the expectation and variance of the number of substitutions K_T at time T is given simply by W_T. Thus, for a constant substitution rate, i.e., for $S(t) = 2k$,

$$E(K_T) = V(K_T) = 2kT. \quad \text{(A1)}$$

For variable substitution rates, i.e., for $S(t) = 2k/(1 + at)$

$$E(K_T) = V(K_T) = [2k \ln(1 + aT)]/a, \text{ with } a > -T^{-1}. \quad \text{(A2)}$$

Expectation and variance of the number of incompatibilities I_T

Expectation and variance of I_T clearly depend on the expectation and variance of K_T and on the kind of incompatibilities being considered, that is epistatic or nonepistatic incompatibilities.

Snowball epistatic incompatibilities Considering a model of Dobzhansky–Muller incompatibilities, where each incompatibility requires exactly two epistatic substitutions, Orr and Turelli (2001) demonstrate

$$E(I_T) = p\mathrm{E}^2(K_T)/2, \quad \text{(A3)}$$

$$V(I_T) = p\mathrm{E}^2(K_T)[1 + 2p\mathrm{E}(K_T)]/2, \quad \text{(A4)}$$

where P is the probability that any given pair of diverged site leads to such an incompatibility. Therefore, considering a constant substitution rate, i.e., $S(t) = 2k$, they found

$$\mathrm{E}(I_T) = 2k^2T^2p, \quad \text{(A5)}$$

$$\mathrm{V}(I_T) = 2k^2T^2p[1 + 4pkT]. \quad \text{(A6)}$$

Equations (A3 and A4) are obviously valid whatever the substitution rate. Expectation and variance of I_T can then easily be obtained using equation (A1) in the case of nonepistatic incompatibilities and variable substitution rates.

Nonepistatic incompatibilities If incompatibilities depend on a single substitution, it is straightforward to show that if K_T follows the distribution given by equation (1), and the numbers of incompatibilities, $I_T = pK_T$ follow a distribution given by

$$p(I_T = n) = (pWT)^n e^{-pWT}/n! \quad \text{where } W_T = \int_0^T S(t')\,dt'. \quad \text{(A7)}$$

The expectation and variance of I_T can then be obtained easily both for constant and variable substitution rates. With a constant substitution rate, i.e., $S(t) = 2kt$, the expectation and variance of the number of incompatibilities after T years of divergence are given by

$$\mathrm{E}(I_T) = V(I_T) = 2kTp. \quad \text{(A8)}$$

Using variable substitution rates, i.e., $S(t) = 2k/(1 + at)$, the expectation and variance of the number of single gene incompatibilities after T years of divergence are given by

$$E(I_T) = \mathrm{V}(I_T) = 2kp\ln(1 + aT)/a \quad \text{with } a > -T^{-1}. \quad \text{(A9)}$$

Expectation and variance of compatibility between species C_T

It is much easier to deduce the expectation and variance of $\ln C_T$ than the moments of C_T directly. Indeed, considering that $\ln C_T = I_T \ln(1 - s)$, where s is the constant deleterious effect per incompatibility (i.e., selection pressure), the expectation and variance of $\ln C_T$ are simply

$$\mathrm{E}(\ln C_T) = \mathrm{E}(I_T)\ln(1 - s), \quad \text{(A10)}$$

$$\mathrm{V}(\ln C_T) = \mathrm{E}(I_T)[\ln(1 - s)]^2. \quad \text{(A11)}$$

Distribution of I_T and C_T

Because the number of substitutions K_T follows a Poisson distribution with mean $W_T = \int_0^T S(t')\,dt'$, the distribution at any one time simplifies to the usual mean and variance $\lambda = 2kT$, assuming a constant substitution rate $S(t) = 2k$. Considering nonepistatic incompatibilities (linear and slowdown models), the number of incompatibilities (I_T) at any time T is simply given by the Poisson distribution with either a constant or variable substitution rate parameter. The Poisson distribution converges eventually on a normal distribution because W_T increases with T. In addition, the probability distribution of the number of incompatibilities follows a Gaussian distribution when substitutions arise at a constant rate and lead to Dobzhansky–Muller incompatibilities (Orr and Turelli 2001). Hence, for all three scenarios we investigate here, the distribution of I_T converges with time to a Gaussian distribution and, accordingly, compatibility C_T converges on a lognormal distribution.

Appendix B

NOTES ON THE DATA

Bacillus data (Zawadzki et al. 1995). Compatibility was estimated via interstrain transformation of *Bacillus* isolates, mostly obtained from the wild. Compatibility values were standardized by reference to within-strain transformation compatibility to give a relative measure with 100% compatibility at 0% DNA divergence. Transformation efficiency was measured by testing the ability of the donor strain to transmit antibiotic resistance to an antibiotic-sensitive strain. Dilution of donor DNA had little effect, but recipient strains from the wild were somewhat variable in transformation probability, and strains with restriction enzymes were particularly slow to transform, presumably because donor DNA was susceptible to restriction enzyme cutting. We excluded such strains from the analysis. The ability to transform in prokaryotes is somewhat similar to eukaryotic prezygotic compatibility, particularly in its dependence on mismatch repair (mismatch repair activity leads to lowered compatibility, as in yeast, below). Mismatch repair strongly enhances sexual isolation in *E. coli*, but has little effect in *Bacillus*; it is thought that sequence divergence in *Bacillus* lowers recombination directly because of a reduction in tendency to form heteroduplex DNA molecules during transformation (Majewski and Cohan 1998). However, as transformation efficiency is measured following survival of the transformed progeny, it may contain elements of "postzygotic" as well as "prezygotic" compatibility. DNA divergence was measured by the authors on a panel of genes (Zawadzki et al. 1995).

Saccharomyces (Liti et al. 2006). Crosses within and between a number of yeasts of the genus *Saccharomyces* were performed, and spore viability was assessed as a measure of compatibility. Recombination requires sequence similarity, and is necessary for successful chromosomal pairing in hybrid *Saccharomyces*, and so diploid hybrids between divergent populations or species tend to be sterile due to meiosis failure. When mismatch repair is

inactivated, fertility improves, suggesting that a direct effect of sequence divergence via its interaction with mismatch repair is the cause of incompatibility (Hunter et al. 1996; Greig et al. 2003). Recombination in both meiotic and mitotic repair declines approximately exponentially with yeast sequence divergence (Chen and Jinks-Robertson 1999); these results suggest that a simple linear or first-order response of mismatch incompatibility to sequence divergence is likely.

Leptasterias starfish (Foltz 1997). Individuals of these starfish were sampled in areas in which a number of cryptic species co-occur. The numbers of F_1 hybrids (identified by allozyme genotypes) sampled in nature divided by the number of individuals of the relevant pure species is the measure of compatibility used here. The hybrid frequencies measured may incorporate postzygotic as well as prezygotic effects on hybrid number.

Alpheus shrimps (Knowlton et al. 1993). Behavioral compatibility of shrimps was measured in a series of experiments in which aggressive and apparently sociable behaviors were scored, and a median compatibility score was constructed ranging from 1 (conspecific compatibility) to zero (all aggressive and no sociable behaviors). Because it refers only to presexual behaviors, and only 1% of heterospecific crosses actually produced fertile egg clutches, these values are likely to be overestimates of overall prezygotic compatibility. In these data, one of the crosses produced a behavioral compatibility value that was higher than the intraspecific values, giving a relative compatibility > 1. Genetic divergence was measured in two different ways: (1) via allozyme divergence (Nei's *D*), and (2) via % mtDNA (*CoI*) divergence.

Drosophila (Coyne and Orr 1989, 1997). Compatibility was estimated from Coyne and Orr's measures of reproductive isolation as 1− (reproductive isolation). Measures of reproductive isolation are of two types. (1) Prezygotic isolation, measured as 1−(frequency of heterospecific matings)/(frequency of within-species matings) in various types of choice or no-choice tests. When the frequency of heterospecific matings was greater than the frequency of homospecific matings (i.e., a negative index was obtained), the index was rounded to 0. (2) For postzygotic isolation a discrete measure was used. If any sex of F_1 hybrid offspring of a single direction of cross between two species *A* and *B* was completely sterile or inviable, reproductive isolation was incremented by 0.25. Reciprocal crosses (i.e., *A* male × B female, versus *B* male × *A* female) may yield different results. Thus, the value for postzygotic isolation varies from 0 (no sex in either reciprocal cross inviable or infertile) to 1 (both males and females sterile or inviable in both directions of cross), but can only take the values 0.00, 0.25, 0.50, 0.75, and 1.00.

Lepidoptera (Presgraves 2002). Genetic distance was based on Nei's *D* value obtained from studies of at least 13 allozyme loci. In the absence of allozyme studies, Nei's *D* was estimated via mtDNA divergence, and converted to Nei's *D* using a regression of Nei's *D* on mtDNA distance (Presgraves 2002). Postzygotic isolation was measured using a method similar to that of Coyne and Orr (1989), and we again estimated compatibility as 1− (reproductive isolation). Presgraves provides two overlapping datasets for postzygotic isolation: hybrid inviability and total postzygotic isolation, both of which we analyze. There were only 13–18 allopatric species pairs for each of these two postzygotic datasets, and as neither Presgraves nor we found any major differences when analyzing sympatric and allopatric species separately, we treated the sympatric and allopatric species together as a single dataset in our analyses. Total reproductive isolation was probably more reliable than the inviability measure, as only four crosses produced completely inviable progeny, and Presgrave's inviability measure averaged only 0.122, whereas average total reproductive isolation was 0.647.

Frogs (Sasa et al. 1998). Various measures of compatibility and isolation were presented by the authors. However, we use only two: (1) a combination of their egg hatch (*EH*) and metamorphosis rate (*MET*), in the form of $EH \times MET$, which gives a compatibility index of survival from egg laying to metamorphosed adult; (2) a measure of compatibility equal to $1 - IPO2$, where $IPO2$ is their discrete measure of postzygotic reproductive isolation. The *EH* and *MET* survival values, and resulting compatibility measure, are continuous measures, but $IPO2$ is an index of hybrid inviability and sterility in discrete units of 0.5 for a single direction of cross, a measure with discrete values 0, 0.5, and 1, somewhat like that used as an isolation index for *Drosophila* (Coyne and Orr 1997). In cases in which there was a measure for $IPO1$ (the average of $IPO2$ for both reciprocal cross directions, with discrete values 0, 0.25, 0.5, 0.75, and 1), we reconstructed the missing $IPO2$ measure so that we were able to use two values of $IPO2$ as suggested in Sasa et al. (1998). For genetic distance, we used the values of Nei's *D* from the same source.

Birds (Price and Bouvier 2002). For our genetic distance measures, we used data for *HKY*-corrected mtDNA divergence, and the Sibley DNA–DNA hybridization measure of $\Delta T_{50}H$. A single measure of compatibility was employed, based on the authors' isolation index. Their "fertility" index, *F*, is like the index of Coyne and Orr, a discrete measure based on the sexes in reciprocal crosses that are inviable or infertile. *F* ranges from 1 to 5 in units of 0.5, with 1 indicating viability and fertility of all hybrids, 5 indicating complete inviability of all hybrids. For our purposes, we also interpret Price and Bouvier's values of $F = 1^*$ to be equivalent to 1.25, and their $F = 5^*$ to be equivalent to 4.5. The compatibility measure we use is then $C = 1 - (F - 1)/4$, which standardizes the measure on a scale of 0 to 1. Thus compatibility may occur in discrete units of 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 0.9375, and 1.000. An essentially complete absence of gene flow is expected from any Price and Bouvier index, $F \geq 3$, equivalent to compatibility $C \leq 0.5$ (all hybrids viable but

infertile); thus our compatibility index overestimates overall hybrid fitness somewhat in the lower compatibility ranges, which may tend to make the fit more snowball-like.

Appendix C

NOTES ON METHODS USED TO FIT THE DATA

Many of the difficulties with least squares fits might be solved by analyzing original datasets, taking into account the sample sizes of each pairwise comparison, using a complete likelihood fitting approach (e.g., GLM). However, the data are highly heterogeneous, making modeling of variation among datasets problematic, and we feel our least squares approach is adequate for the broad exploratory overview attempted here.

Another potential problem is nonindependence of comparisons of the same or related species. Traditionally, this has been resolved by taking into account phylogenetic relatedness to obtain independent contrasts of phylogenetic nodes (e.g., Coyne and Orr 1997; Sasa et al. 1998; Presgraves 2002; Bolnick and Near 2005). However, although this approach is valuable for avoiding incorrect rejections of null hypotheses, it is not nearly so helpful for curve-fitting and parameter estimation. In no case so far analyzed have phylogenetic corrections made a difference to major conclusions (Coyne and Orr 1997; Price and Bouvier 2002; Mendelson et al. 2004). Although compatibility data on crosses among three species (say $A \times B$, $A \times C$ and $B \times C$) almost certainly contain some information that is not independent, it may also contain much that is independent, particularly if different laboratories have performed different experiments. "Asymmetric" postmating isolation is extremely common (Turelli and Moyle 2007), suggesting a certain amount of independence even between reciprocal directions of the same cross, let alone between related crosses. Methods used to correct for phylogenetic nonindependence in incompatibility data have varied among studies (Coyne and Orr 1997; Bolnick and Near 2005), and indeed there is no clear way to decide which method is best (Bolnick and Near 2005). A similar nonindependence problem affects the argument whether Haldane's Rule (i.e., the bias toward the heterogametic sex in unisexual incompatibility) is significant. In one scenario, every speciation event is imagined to be independent; in another, the tendency to obey the rule by all species in each major group with a particular kind of heterogamety (*XY* vs. *ZW*, e.g., *Drosophila*, birds, Lepidoptera, mammals) might be due to a major phylogenetic correlation (Read and Nee 1991). Therefore, in this exploratory analysis we prefer to acknowledge that *P*-values from our *F*-ratio tests are unreliable, and to draw attention to conclusions that might be dubious because of phylogenetic correlations. Appropriate corrections will anyway mainly act to reduce the effective degrees of freedom (i.e., sample size) in a fit; previous work has suggested that this reduction will be typically of the order of 20%–50% (Coyne and Orr 1997; Fitzpatrick 2002; Presgraves 2002; Bolnick and Near 2005).

Genetic distances between species are used here as surrogates for time since divergence, as in most previous work (Coyne and Orr 1997; Sasa et al. 1998; Presgraves 2002; Price and Bouvier 2002). We should be cautious about this, due to variation in rates of molecular evolution (Bolnick and Near 2005). There is some evidence for a better fit of *Drosophila* reproductive isolation data with Nei's *D* based on allozymes (which depends on differences in amino acid sequence) than with % divergence based on mtDNA (Fitzpatrick 2002). In that study, selection was argued to be responsible for divergence of allozymes as well as reproductive isolation, hence the better fit. However, Fitzpatrick's results do not seem altogether relevant for our purposes for a number of reasons: (1) Fitzpatrick's analyses were performed only via simple linear regressions based on "optimal" transformations of both axes, rather than using constrained model-fitting approaches as here. For example, reproductive isolation evolution was allowed to be unconstrained, and did not even force isolation to be zero given zero genetic distance. (2) There is no obvious reason why enzyme divergence itself should affect reproductive isolation directly. (3) It is not obvious why the same processes that affect variation in levels of positive selection on allozymes should affect reproductive isolation at the same time and in the same way; and (4) the tighter dependence on allozyme divergence might be due primarily to weaker molecular clock information and greater variability in rate from the single-locus mtDNA sequence data used rather than a real effect of selection on reproductive isolation. In summary, genetic distance is a surrogate, but it generally correlates reasonably well, although noisily, with time since divergence. Genetic distance samples generalized rates of substitution, and so should be close to the kind of distance measure we need to estimate the probability of neutral substitution across the genome that could lead to incompatibility. Although we normally think of our analysis as a test of models of compatibility decline over time, more strictly it tests these models relative to overall levels of substitution at the loci used in measuring distance.