Evolution, 59(8), 2005, pp. 1620–1632

COMPENSATORY EVOLUTION OF INTERACTING GENE PRODUCTS THROUGH MULTIFUNCTIONAL INTERMEDIATES

ERIC S. HAAG[1,2] AND MICHAEL N. MOLLA[3,4,5]
[1]*Department of Biology, Building 144, University of Maryland, College Park, Maryland 20742*
[2]*E-mail: ehaag@umd.edu*
[3]*Department of Computer Science, University of Wisconsin, Madison, Wisconsin 53704*
[4]*E-mail: molla@cs.wisc.edu*
[5]*Living Circuit, LLC, 412 South Brearly Street, Number 1, Madison, Wisconsin 53703*

Abstract.—When two mutations are singly deleterious but neutral or beneficial together, compensatory evolution can occur. The accumulation of derived, compensated genotypes contributes to the evolution of genetic incompatibilities between diverging populations or species. Previous two locus/two allele models have shown that compensatory evolution is appreciable only with tight linkage, the possibility of nearly simultaneous mutations, and/or a way to overcome negative selection against the singly mutated genotype. These conditions are often not met. Even when they are met, compensatory evolution is still predicted to be extremely slow, and in many scenarios selective advantage of the compensated genotype does little to accelerate it. Despite these obstacles, empirical studies suggest that it occurs readily. We describe here a set of related two locus/three allele models that invoke plausible neutral intermediates capable of productive interaction with both ancestral and compensated products of the interacting locus. These models are explored with analytical and computer simulation methods. The effect of these stepping-stone alleles on the evolution of ancestor-descendant incompatibilities is often profound, making the difference between evolution and stasis in several situations, including in small populations, when codominance or haploidy prevents shielding of mismatched genotypes, and in the absence of positive selection on the derived genotype. However, in large populations these intermediates can either speed or slow the evolution of incompatible genotypes relative to the two-allele case, depending on the specific fitness model. These results suggest that population size, the source of adaptive benefit, and the structural details of heteromeric gene product complexes interact to influence the path by which intergenic incompatibility evolves.

Key words.—Coevolution, epistasis, heterodimer, incompatibility, simulation.

Received September 7, 2004. Accepted May 18, 2005.

Interactions between gene products are crucial to the function of cells and organisms. For example, cell adhesion, signal transduction, gene regulation, fertilization, and immune function are all mediated by highly specific protein-protein or protein–nucleic acid complexes. Given the complexity of such interactions, significant evolutionary divergence would seem to be difficult. However, the genes encoding interacting products evolve—in some cases with remarkable rapidity. Reproduction-related protein interactions are especially extreme examples (e.g., Swanson and Vacquier 1995; Metz and Palumbi 1996; Hellberg and Vacquier 1999; Kachroo et al. 2001; Haag et al. 2002). These changes may simply provide equally useful alternate solutions to a common biochemical need and thus be neutral overall. Others may represent important adaptations, as may be the case when a derived genotype allows for a novel intermolecular interaction, or for changes in the specificity of an existing interaction. In either case, an important feature of evolving interactions is epistasis—the functionality of a complex is necessarily a synthetic feature of multiple gene products.

These synthetically neutral or adaptive changes also produce incompatibilities between the ancestral and descendant genotypes. At its simplest, such incompatibility can be represented by the shift of alleles at two interacting loci, A and B, to another pair of alleles, A′ and B′, which can interact productively with each other but not in combination with A or B. Barring simultaneous mutations, evolution from AB to A′B′ would seem to require individual mutations that produce dysfunctional intermediates, often described as a ''fitness valley between adaptive peaks.'' Crossing the valley requires a second, compensatory mutation to occur before selection can drive the first extinct, a process recently described as ''stochastic tunneling'' (Iwasa et al. 2004). Models explored early on by Haldane (1931) and Wright (1931, 1932; see elaboration in Kimura 1990) indicated that compensatory changes in interacting genes could indeed occur. More recent work (e.g., Crow and Kimura 1965; Gillespie 1984; Kimura 1985; Barton 1989; Michalakis and Slatkin 1996; Phillips 1996; Stephan 1996; Carter and Wagner 2002), however, has shown that such change is only expected to be appreciable with tight linkage of the interacting loci or sites, high mutation rates, and/or very weak selection against intermediates.

Despite these theoretical results, compensatory evolution occurs in essential gene products encoded by both linked and unlinked loci. Reproduction-related examples include proteins mediating self-incompatibility in cruciferous plants (Kachroo et al. 2001), species-specific gamete interactions in marine invertebrates (Swanson and Vacquier 2002), and the signal transduction pathway mediating nematode sex determination (Haag et al. 2002). Recently, pathogenic human mutations fixed in mouse orthologues (so-called compensated pathogenic deviations) have been inferred to represent examples of either intra- or intermolecular compensation (Kondrashov et al. 2002). These examples suggest that, in practice, compensatory evolution is not as difficult as current theory would lead us to believe. Are we missing something? Gavrilets (1997) noted that whenever there are high-fitness genotypes that allow a way around a valley, a ''holey landscape'' is a better metaphor than one composed of peaks and valleys. Indeed, for linked sites of ribosomal RNA, the G:U

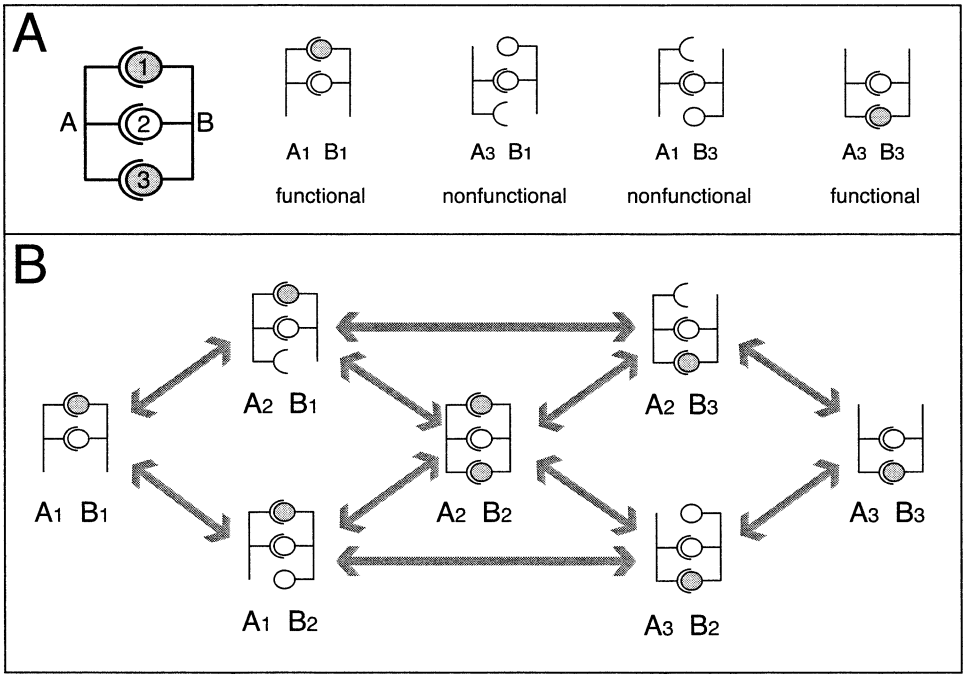


FIG. 1. Compensatory versus pseudocompensatory evolution of interacting gene products. The products of two genes, A and B, interact via multiple bonds, numbered here 1–3. Any one of these bonds is dispensable. In both true compensatory evolution and in pseudocompensation, ancestral alleles allowing interaction via bond 1 could evolve into descendant alleles functioning via bond 3 (both indicated by the shaded ovals). Bond 2 does not change, but is included to represent the multipartite nature of macromolecular interactions. (A) In traditional compensatory evolution, fully functional allele pairs, such as A_1B_1 or A_3B_3, are separated mutationally by nonfunctional mismatch genotypes that have only a single bond. With strong selection against mismatched alleles, evolution from A_1B_1 to A_3B_3 is possible only under a narrow range of conditions. (B) In pseudocompensation, a multifunctional adaptor allele at each interacting locus (A_2 or B_2) is posited to exist that has the capacity to interact productively with the products of both ancestral and derived alleles. These enable a set of allelic substitution pathways from A_1B_1 to A_3B_3 in which all intermediate genotypes are fully functional, as defined by possessing at least one of the variable bonds (indicated with shading). Although back-mutation (indicated by left-pointing arrowheads) is possible, we do not formally consider it here.

base pair has been shown to be just such an intermediate, allowing facile transitions between canonical base pairs (Rousset et al. 1991).

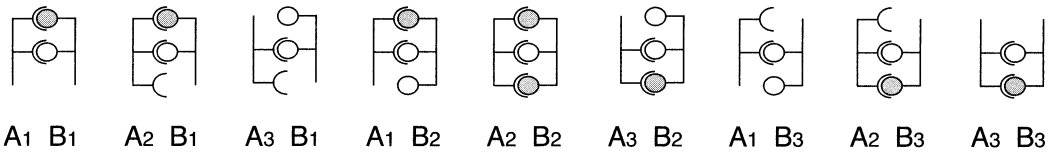
In this spirit, we wondered if the interacting products of unlinked genes might perhaps avoid the ''valley of mismatched alleles'' via intermediate forms that possess excess capacity (i.e., forms capable of interacting productively with the products of multiple partner alleles). In this case, what appears to be compensatory change does not actually involve correction of an individually deleterious mutation. We dub this phenomenon ''pseudocompensation'' for clarity. Using computer simulations, we find that in many cases the existence of such intermediates can indeed enable the evolution of ancestor-descendant incompatibilities where it could not otherwise occur. However, in large diploid populations these same intermediates can either speed or hinder adaptive evolution, depending on the extent of dominance and to a surprising extent upon the exact structural basis for the benefit of the pseudocompensated genotype.

MODELS, PREDICTIONS, AND SIMULATION METHODS

Models

Our model is based on a simple scenario for the evolution of intermolecular incompatibilities that can accommodate intermediates. Two gene products, A and B, are posited to require at least two points of contact to (henceforth ''bonds'') to support a fully productive interaction (Fig. 1). Starting with a population whose A-B interaction is mediated by bonds 1 and 2 (alleles A_1 and B_1), incompatibility between ancestor and descendant occurs when the descendant population evolves an A-B interaction dependent upon the ancestral bond 2 and the novel bond 3 (alleles A_3 and B_3). Traditional models of compensatory evolution only consider the ancestral and descendant alleles and necessitate the existence of mismatched transitional genotypes (Fig. 1A).

In pseudocompensation (Fig. 1B), intermediate alleles A_2 and B_2 are posited to exist such that both have the potential to form all three bonds. They interact equally well with ancestral and derived alleles, although in each case an unused half-bond will exist in the A_2 and B_2 alleles. We assume here that this unused capacity will confer no selective disadvantage. Since bond 2 is present in all possible allelic combinations, its presence is actually not essential for implementation of the model. We include it to emphasize that dimerization is generally dependent upon the cumulative effect of many bonds and to clarify why mismatched alleles may still have considerable interaction potential. If all three bonds were allowed to vary, there are three different two-bond allele pairs: the two discussed in the previous paragraph plus one with bonds 1 and 3 only. Our scenario could therefore also be thought of as modeling one of the six possible transitions



	A_1B_1	A_2B_1	A_3B_1	A_1B_2	A_2B_2	A_3B_2	A_1B_3	A_2B_3	A_3B_3
s_A for absence of bond 1	1	1	$1-s_M$	1	1	$1+s_A$	$1-s_M$	$1+s_A$	$1+s_A$
s_A for presence of bond 3	1	1	$1-s_M$	1	$1+s_A$	$1+s_A$	$1-s_M$	$1+s_A$	$1+s_A$
$s_A/2$ for absence of bond 1; $s_A/2$ for presence of bond 3 (additive)	1	1	$1-s_M$	1	$1+s_A/2$	$1+s_A$	$1-s_M$	$1+s_A$	$1+s_A$

FIG. 2. Haploid epistatic fitness models of interacting macromolecules. The possible combinations of the three allelic products of gene A and gene B are diagrammed along the top. Each row below the diagrams shows the corresponding fitnesses when the fully incompatible (A_3B_3) or semi-incompatible (A_2B_3, A_3B_2) genotypes have a fitness higher than the ancestral genotype, A_1B_1, by the amount s_A. In the first row, s_A is awarded for any successfully interacting pair that lacks a complete ancestral bond 1. In the second row, s_A is awarded for the presence of the derived bond 3. Finally, the third row represents the additive case, where each criterion counts for $s_A/2$. Mismatched alleles, in which neither bond 1 nor bond 3 is present, pay a fitness penalty, s_M.

between any two of them, all of which would behave with the exact same dynamics.

Although dysfunctional allelic combinations can be avoided in pseudocompensation, more mutational events are required to reach a state of ancestor-descendant incompatibility. A formal model is therefore needed to explore the conditions under which intermediate alleles facilitate the evolution of ancestor-descendant incompatibility. In pseudocompensation via intermediates, the exact fitness of a given haploid genotype depends on the basis for any selective advantage, s_A, that may exist for a derived allele pair (Fig. 2). One possibility is that the derived pair is better because it has ceased to use one of the two ancestral bonds, which is constrained here to be bond 1 for simplicity. This situation might apply where reduction of pleiotropic interactions or an increase in molecular crypsis is important. Alternatively, the gain of bond 3 may alone confer advantage. Finally, both may be important, in which case we consider their effects to be additive. Despite these differences, only the fitness of the A_2-B_2 pair varies with the structural basis of selective benefit. In all three scenarios the A_2-B_3, A_3-B_2, and A_3-B_3 allele pairs all have the same fitness, $1 + s_A$.

TABLE 1. Diploid fitness models used in the simulations.

Fitness model	Structural basis for adaptation	Diploid implementation
DM1	s_A for absence of bond 1	best pair dominant
DM2	s_A for presence of bond 3	best pair dominant
DM3	$s_A/2$ for absence of bond 1 $s_A/2$ for presence of bond 3	best pair dominant
DM4	s_A for absence of bond 1	codominant (greedy)
DM5	s_A for presence of bond 3	codominant (greedy)
DM6	$s_A/2$ for absence of bond 1 $s_A/2$ for presence of bond 3	codominant (greedy)

Table 1 illustrates six ways in which these three scenarios can be used to determine the fitness of diploid organisms. Three diploid fitness models (DM1, DM2, and DM3) use the best pair dominant (BPD) scenario, in which only a single allele pair is required for full function, and the one with highest fitness is chosen. This scenario is justified by the recessivity of most loss-of-function mutations and by the principle that the kinetics of binding will be fastest for well-matched gene products. Less-matched products will simply be redundant or unproductive in a way that causes no harm. This should be biologically realistic for neutral pseudocompensatory evolution (i.e., when $s_A = 0$), where all fully interacting combinations are equally fit, and in adaptive cases where the novelty of new allele pairs is more important for adaptation than the removal of ancestral alleles. A codominant scenario is also considered (DM4, DM5, and DM6), in which the fitnesses of the two allele pairs are averaged. A ''greedy'' algorithm is used here, whereby the highest fitness

pair is matched first and the remaining two alleles are forced together. In the two-allele case, this scenario is identical to the additive metabolic model of Phillips (1996) and applies in cases where the function of the gene products is dose dependent or where specificity is especially important for adaptation. In cases where two different best pairs could be formed, the one that produces the highest total fitness is chosen.

In the analysis and simulation runs described below, the entire population starts with the A_1B_1 (haploid) or the $A_1A_1B_1B_1$ (diploid) genotype. This monomorphic starting point is justified by the observation that most macromolecular residues that differ between species are not detectably polymorphic within species. Variable parameters are the population size, number of generations, degree of selection against mismatched allele pairs (s_M), selective advantage for compensated allele pairs (s_A), and the six allowed forward mutation rates (A_1 to A_2, A_1 to A_3, A_2 to A_3, B_1 to B_2, B_1 to B_3, and B_2 to B_3). We note that if each half-bond is thought of as a modular mutable element, such as an amino acid, then the A_1-to-A_3 and B_1-to-B_3 transitions would seem to require two mutations and thus occur at the square of the frequency of the others. However, mutations that cause conformational changes could simultaneously create half-bond 3 and eliminate half-bond 1. Without a rational basis to estimate the probabilities of these two options, we have elected to keep all nonzero mutation rates equal. Back-mutation is not allowed. When neutral compensation is considered ($s_A = 0$) or when the intermediates are not allowed (by setting all but the A_1-to-A_3 and B_1-to-B_3 mutation rates to zero), these six diploid models collapse to two, one BPD and one codominant.

An exact analytical treatment of pseudocompensation that takes both polymorphism and the various diploid fitness models into account is extremely difficult. However, some sort of predictive framework is useful to aid in the interpretation of the simulation results that follow. For this purpose, we develop below a greatly simplified model of pseudocompensation, in which it is viewed as a series of either neutral or adaptive allelic fixations (Fig. 3). As shown by Kimura (1983), in diploid populations the rate of fixation for new neutral mutations is the mutation rate, μ, and the rate for adaptive mutations is $4Ns\mu$, where N is the population size and s is the positive selection coefficient. Because our simulations allow two symmetrical mutational paths for each step (for example, the $A_1A_1B_1B_1$ genotype can change with equal probability to $A_2A_2B_1B_1$ or $A_1A_1B_2B_2$), the total network (Fig. 1B) can be collapsed into two linked pathways (direct vs. indirect) in which neutral steps occur with rate 2μ and adaptive steps with rate $8Ns\mu$ or $4Ns\mu$. The four scenarios in Figure 3A depict the awarding of s_A for bond loss (approximating DM1), the awarding of s_A for bond gain (approximating DM2), additively awarding $s_A/2$ for both bond loss and bond gain (approximating DM3), and when pseudocompensation is neutral.

Because the 2-type intermediate alleles do not necessarily need to reach fixation for the subsequently arising 3-type alleles to gain traction, this simplification will underestimate the speed of evolution. However, our simulation results (data not shown) indicate that sequential fixation occurs in all adaptive runs where $N = 100$, and about half the time when $N = 1000$. It occurs in all population sizes when pseudocompensation is neutral. Thus, this underestimate of rate primarily applies to adaptive evolution in large populations. The analysis below uses the sequential fixation approximation to predict the role of various parameters on the evolution of incompatibility over time. In the simulations that follow the analysis, we allow cohorts of 20 replicate populations to evolve from the ancestral genotype toward incompatibility with the ancestor under a variety of conditions.

Adaptive Bond Loss Scenario

We first consider the scenario where s_A is awarded for loss of an ancestral bond. In all four scenarios (Fig. 3A), there is both a direct path, in which an ancestral allele (A_1 or B_1) mutates directly to a derived allele (A_3 or B_3) after the neutral fixation of an intermediate allele (B_2 or A_2) at the interacting locus, and an indirect path, in which both intermediate alleles fix in succession, followed by mutation of one gene to the derived allele. Because the indirect path requires an additional mutational step, the tendency for populations to use this path should be important for both the overall rate and mean time of pseudocompensation.

Let d be the fraction of populations taking direct path, and i be the fraction of populations taking indirect path. d and i are estimated as

$$d = \frac{8Ns_A\mu}{8Ns_A\mu + 2\mu} = \frac{4Ns_A}{4Ns_A + 1} \quad \text{and} \tag{1a}$$

$$i = \frac{2\mu}{8Ns_A\mu + 2\mu} = \frac{1}{4Ns_A + 1}. \tag{1b}$$

When the product $4Ns$ is fairly large (i.e., 10 or more), as is the case in these simulations, d approaches unity and the fraction taking the indirect path will be small. Combined with the extra mutational step required for the indirect path, it will contribute little to compensatory change and can safely be ignored.

Having established that the direct path will be almost exclusively used in the bond-loss scenario, we can now estimate the overall rate of pseudocompensatory evolution from the ancestral genotype to the semi-incompatible state ($A_2A_2B_3B_3$ or $A_3A_3B_2B_2$). By analogy with sequential first-order chemical reactions, the conversion of populations taking the direct path to the semi-incompatible genotype, p_d, for any time, t, is given by

$$p_d = 1 + \frac{k_2^{-k_1 t} - k_1^{-k_2 t}}{k_1 - k_2}, \tag{2}$$

where k_1 and k_2 are the rates of the first and second transitions, respectively. The above general form describes adaptive bond loss when $k_1 = 2\mu$ and $k_2 = 8Ns_A\mu$. Assuming N is reasonably large, the overall fraction of populations having reached either $A_3A_3B_2B_2$ or $A_2A_2B_3B_3$ from $A_1A_1B_1B_1$, p_ψ, at any time, t, is thus given by

$$p_\psi \approx 1 + \frac{2\mu^{-(8Ns_A\mu)t} - 8Ns_A\mu^{-(2\mu)t}}{8Ns_A\mu - 2\mu}. \tag{3}$$

Figure 3B shows a plot of equation (3) for the mutation

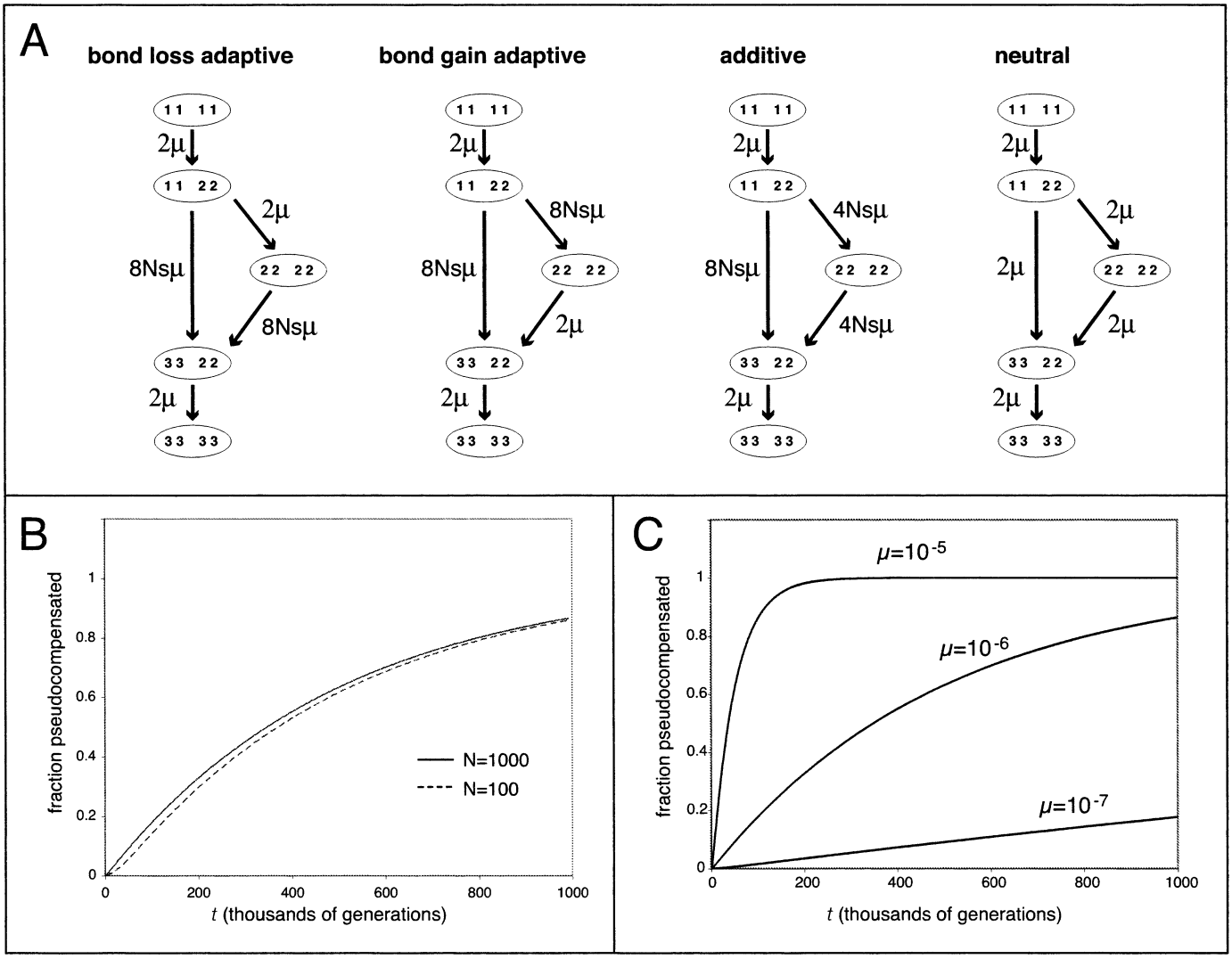


FIG. 3. Simplified schemes of pseudocompensation. (A) The six possible unique mutational pathways, being symmetrical, are collapsed into two—direct and indirect. Each oval represents the diploid genotype for the two interacting loci. For example, ''1122'' represents both $A_1A_1B_2B_2$ and $A_2A_2B_1B_1$. For each fitness scenario, the multiple allelic substitutions are modeled as series of sequential neutral or adaptive allelic fixations, with the probability of each transition indicated along the path between genotypes. (B) Predicted cumulative evolution of the semi-incompatible genotypes ($A_2A_2B_3B_3$ or $A_3A_3B_2B_2$) under the adaptive bond loss scenario with the assumption of sequential fixation and a mutation rate, μ, of 10^{-6}. (C) Dependence of the rate of pseudocompensation on mutation rate, μ, under the adaptive bond loss scenario and a population size of $N = 1000$.

rate ($\mu = 10^{-6}$), positive selection coefficient ($s_A = 0.05$), and two population sizes ($N = 100, 1000$) used in the simulations. Population size has a relatively minor effect because it only affects the second, adaptive rate. Thus, when the quantity Ns_A is much greater than one, formation of the intermediate allele by mutation and drift is rate limiting. Population sizes larger than 1000 do not differ appreciably from each other in this specific example. However, since polymorphism is not considered here, equation (3) will give increasingly erroneous estimates as population size increases. The simulations below show that the magnitude of this error varies greatly with the diploid fitness model. Figure 3C shows that equation (3), and thus pseudocompensation, is extremely sensitive to the mutation rate. Since N is a factor in the rates of both adaptive and neutral steps, this is not surprising.

Adaptive Bond Gain Scenario

When s_A is given for the gain of a new bond or when pseudocompensation is neutral, the two possible forward steps from the $A_1A_1B_2B_2$ or $A_2A_2B_1B_1$ genotypes occur at the same rate (Fig. 3A). Thus d and i are both 1/2, and the usage of the indirect path to the overall rate of pseudocompensation must be considered. Given that both paths pass through the common $A_2A_2B_1B_1$ or $A_1A_1B_2B_2$ intermediate, the relative contribution of each pathway at any given time can be found by evaluating the functions describing direct versus indirect formation of the semi-incompatible genotype from this intermediate.

Using conditions like those in the simulations that follow, the direct path is much faster than the indirect path because

the latter imposes a much slower, rate-limiting neutral step. For example, when $N = 1{,}000$, $s_A = 0.05$, and $\mu = 10^{-6}$, by 20,000 generations all populations taking the direct route have reached the semi-incompatible state, but only 3% of those taking the indirect path have. Thus, as with the bond-loss scenario, the indirect path in the bond-gain scenario can largely be ignored in calculating the maximal rate of pseudocompensation. However, because only half of the populations take the fast, direct route this rate will be approximately half of that in the bond-loss case. Because the simulations presented below track the evolution of cohorts of populations until all of them reach at least the semi-incompatible state, we also note that the use of the indirect route by half of the populations will create a biphasic accumulation of pseudocompensated populations and greatly increase the mean time to semi-incompatibility of the cohort as a whole.

Additive Scenario

In the additive scenario, $s_A/2$ is given for gain of bond 3 and $s_A/2$ for the loss of bond 1. As a result, the indirect mutational path differs from that of the previous two cases in that it has two adaptive phases and no rate-limiting neutral step (Fig. 3A). In small to moderate sized populations for which the sequential fixation approximation holds, the fraction of populations taking the direct path should be about twice that using the indirect path, though their rates will be similar. However, as will become clear from the simulations below, the most interesting feature of the additive model is that it allows rapid pseudocompensation in haploid or codominant conditions where the direct route is disfavored due to selection against even rare 3-type alleles.

Neutral Pseudocompensation

For neutral pseudocompensation the rates of all fixations are 2μ (as the indirect path function is undefined when $k_1 = k_2$, numerical evaluation requires tiny deviations in k_1 and k_2). In this case, the indirect path lacks a slower rate-limiting step relative to the direct path, and thus contributes nearly half to the rate neutral pseudocompensation. This rate is expected to be sensitive only to mutation rate and not to population size or structural fitness model, and will therefore not be considered here further. However, we present simulations of neutral pseudocompensation below for comparison with the adaptive cases.

Method for Simulations

The first step in each simulation is to calculate the relative fitness of each of the nine (haploid) or 81 (diploid) possible genotypes given s_M, s_A, and the particular fitness scenario specified. Each generation of the simulation consists of three phases: mutation, survival, and mating. In the mutation phase, each individual in the population is mutated with probability described by the given six forward mutation rates. The resulting population is then subjected to the survival phase in which the quantity of each of the possible genotypes is adjusted directly by its precomputed relative fitness. Finally, in the mating phase, a number of gametes equal to twice the population size are formed by pulling random pairs of A and B alleles (with replacement) from the population pool, and these are then randomly paired to form the next generation. The mutation, survival, and mating phases are then repeated. Note that, though the number of individuals in the population may diverge from the given population size during the survival phase, it will return to its original size during the mating phase. This ensures a population of consistent size at the start of each iteration.

The C++ program that performs the haploid simulations, SYNTH_POP_HAP, was compiled and run on a Sun cluster at the University of Wisconsin, Madison. The diploid version, SYNTH_POP, is much slower, and was compiled on a Linux cluster to run up to 50 simultaneous runs with the gracious assistance of Kai Zhang and the University of Maryland Institute for Advanced Computer Studies. Both sets of simulations were run using the Condor Software Program, developed by the Condor Team at the Computer Sciences Department of the University of Wisconsin, Madison (available at http://www.cs.wisc.edu/condor/downloads/v6.6.license.html). For each of the 30 unique combinations of selection strength, fitness model, and population size tested, 20 replicate runs were performed using the same 20 random seeds (1–20). To avoid potential artifacts due to random number generator periodicity, the seeds of simulations that had not reached at least semi-incompatibility by the 10 millionth generation were changed by adding a ''1'' in front of each (e.g., 13 becomes 113). This was only required for two runs of the 840 performed, both in the haploid adaptive bond gain simulations with $N = 10{,}000$ (seeds 15 and 19). Depending on population size, some diploid runs took as long as several weeks to complete. Data are output as tab-delimited files that note the frequency of each of the possible ordered genotypes at intervals of 200 generations. The source codes for SYNTH_POP and SYNTH_POP_HAP are available from the authors on request, or they can be downloaded from http://www.cs.wisc.edu/~molla/synth_pop.

RESULTS

Simulating the Two-Allele Case

To verify that the models could recapitulate the previously described barriers to truly compensatory evolution in the two-allele case, adaptive ($s_A = 0.05$) runs without intermediates and with strong selection against mismatched allele pairs ($s_M = 0.05$) were performed for three population sizes (100, 1000, and 10,000; Fig. 4). In these and all subsequent runs, all nonzero mutation rates were 10^{-6} (which we justify as representing a set of 100 mutable sites with an absolute per nucleotide mutation rate of 10^{-8}). For the haploid and codominant scenarios, compensated changes never evolve because mismatched A_3 or B_3 alleles are strongly selected against at all frequencies. In the BPD diploid model, evolution of the compensated genotype occurs in a strongly population size-dependent manner: after 1 million generations compensation occurred 100% of the time when $N = 10{,}000$, 15% when $N = 1000$, and 0% when $N = 100$. This pattern can be explained by the capacity of larger populations to harbor significant numbers of singly heterozygous individuals when the mismatched 3-type allele is recessive (Haldane 1932; Wilson and Bossert 1971). The size of this subset of

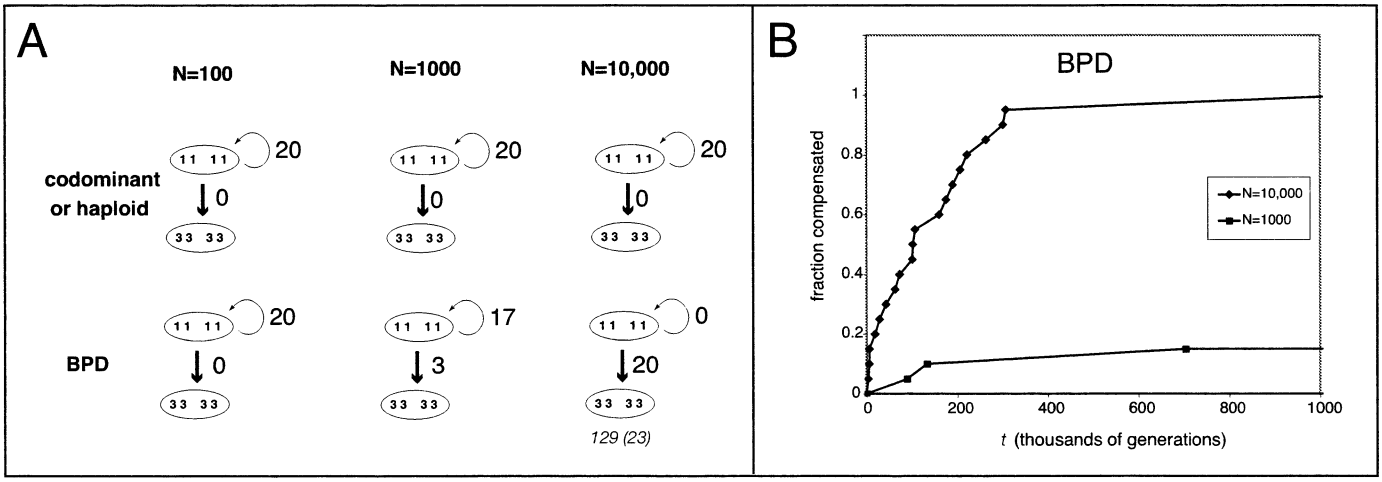


FIG. 4. Simulation of adaptive compensatory evolution in the two-allele model. (A) Each diagram represents the summary of 20 independent simulations of evolution of the starting population fixed for the genotype $A_1A_1B_1B_1$ (A_1B_1 in haploids), under the indicated fitness models and population sizes. Ovals with numbers represent genotypes, as described in Figure 3. Arrows circling back to an oval represent cases where the indicated number of populations fixed at that genotype at the end of 1 million generations. Arrows between the ovals represent transitions from the ancestral genotype to the compensated genotype, with the number of populations making the transition within the first million generations indicated. The number below the diagram in the lower right is the mean number of generations (in thousands) required to reach fixation of A_3 and B_3, with the standard error in parentheses. In all cases, $s_M = 0.05$ and $s_A = 0.05$, and the mutation rates are 10^{-6}. BPD, best pair dominant. (B) Accumulation of populations fixed for the compensated, incompatible genotype ($A_3A_3B_3B_3$) under three populations sizes (100, 1000, and 10,000). The line for $N = 100$ coincides with the x-axis.

single heterozygotes is directly related to the probability of a compensatory mutation occurring in the other locus. However, even in the most favorable cases the mean time to the evolution of a compensated genotype was 129,000 (SE = 23,000) generations, confirming that compensatory evolution is generally slow or nonexistent in unlinked genes.

The Effect of Intermediate Alleles When Compensatory Change Is Adaptive

Because of the identical fitness of the semi-incompatible allele pairs A_2B_3 and A_3B_2 with the fully incompatible A_3B_3, A_3B_3 must be reached by drift alone. This occurred in about half of the various diploid simulation runs before 1 million generations, but in less than a quarter of the haploid simulations. We therefore use the timing of the evolution of the semi-incompatible genotypes to compare the rates of compensation versus pseudocompensation in the two- and three-allele models, respectively. To determine mean times to incompatibility, individual seeds were rerun as needed to allow them to reach at least a semi-incompatible genotype.

As described above, in the absence of intermediate alleles only large populations with strong dominance of the fittest allelic combination consistently evolve adaptive, ancestor-incompatible genotypes. When the possibility of mutation to the A_2 and B_2 intermediate alleles is added to the simulations, however, semi-incompatibility (and often complete incompatibility) typically evolves in less than 1 million generations, even in small populations and with codominant or haploid fitness models (Figs. 5, 6). This demonstrates that the intermediate alleles create a fitness ridge to an adaptive genotype that is used even though it requires more mutational steps. Less intuitive, though, are the major differences in kinetics imposed by population size and the different diploid fitness models.

In the above analytical treatment of an adaptive bond-loss scenario with small ($N = 100$) to medium ($N = 1000$) population sizes and a global mutation rate of 10^{-6}, semi-incompatibility was predicted to evolve in 50% of populations in roughly 400,000 generations and in over 80% of populations after 1 million generations. The adaptive bond-loss simulations using DM1 and DM4 (Fig. 5, upper left panel) are in good agreement with this, and show that its BPD implementation, DM1, is overall the fastest-evolving model. These two models differ primarily in the extent to which evolution is accelerated in the largest population size ($N = 10{,}000$). As expected, in DM1 pseudocompensation in large populations is greatly enhanced due to the fact that the first, neutral intermediate allele is far from fixation when the second, synthetically adaptive A_3 or B_3 allele arises and begins to increase. This occurs due to the reinforcing effects of two features. One is the ability of BPD models to shelter rare A_3 or B_3 alleles from selection against mismatched allele pairs, but still allow them selective advantage when paired with a neutral intermediate at the other locus. In this sense, the BPD model allows the A_3 and B_3 alleles ''to have their cake and eat it too.'' The second feature of DM1 that makes it particularly sensitive to population size is the exclusive use of the direct mutational pathway in large populations (see substitution pathway diagrams in Fig. 5). As discussed above and illustrated in Figure 3A, this allows populations to avoid an additional rate-limiting step that is independent of population size.

Population size dependence is also strong in the additive BPD model, DM3, which evolves to semi-incompatibility in large populations at a rate not significantly different from DM1. The simulations show that the speed of large populations under DM3, as with DM1, is due to exclusive use of the direct mutational path. In the model of successive fixation

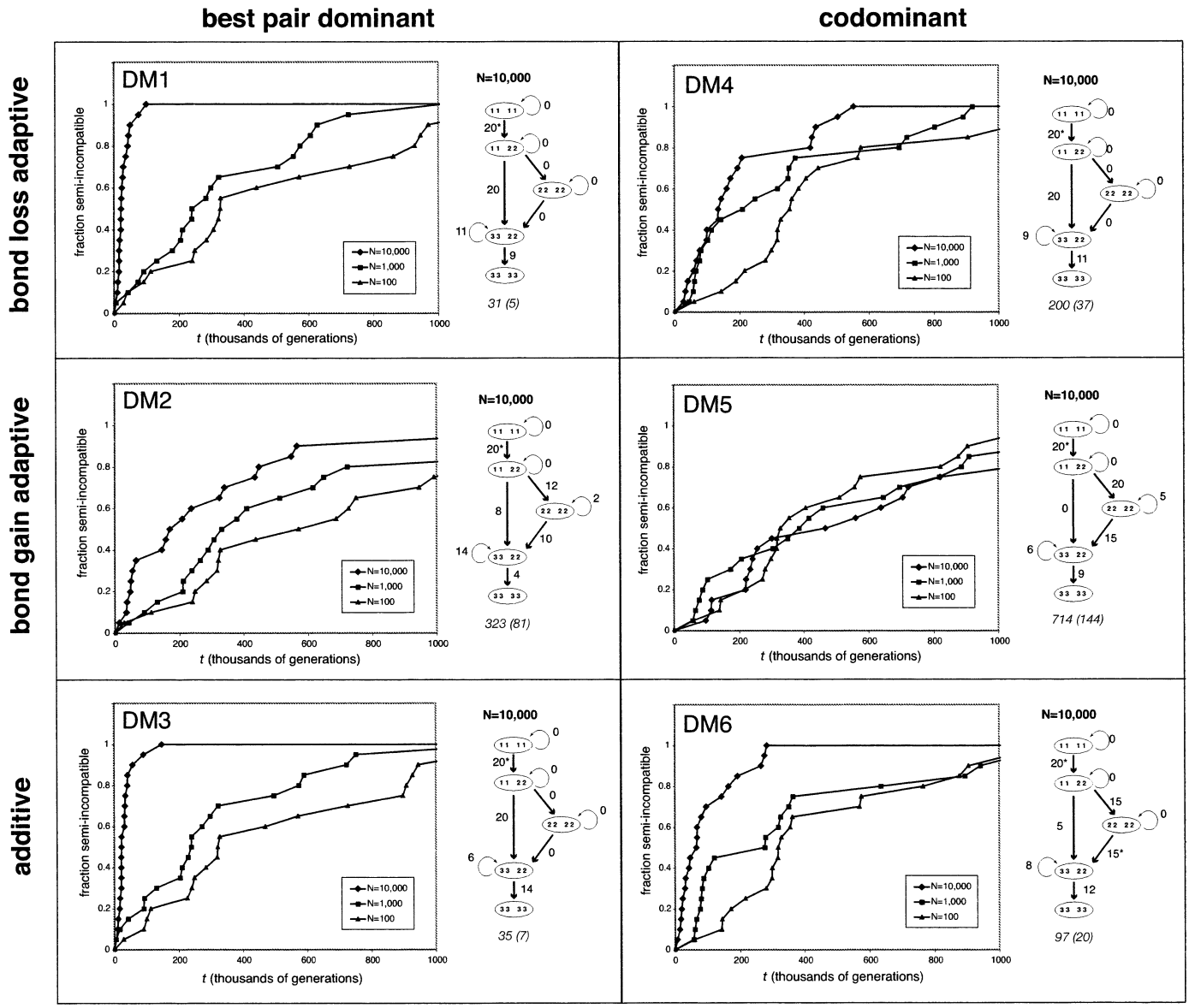


FIG. 5. Diploid adaptive pseudocompensation. Twenty simulation runs (each with a unique seed number) were generated per model for each of three population sizes. The left portion of each panel shows the accumulation of populations reaching the semi-incompatible state over the first million generations. The allelic substitution diagrams for the first million generations, shown on the right of each panel, are as described in Figure 4 (with the addition of intermediate genotypes), and the fitness models DM1–DM6 are explained in Table 1. The asterisks after the number of populations taking a mutational step indicate that in most or all of these simulations the target genotype failed to reach fixation before the subsequent genotype began to increase. Below each substitution diagram is the mean time to semi-incompatibility (in thousands of generations) for all 20 populations. This was obtained by extending the run time as needed for those seeds that failed to reach a semi-incompatible genotype in the first million generations. Similarities in $N = 100$ and $N = 1000$ curves under different fitness models represent identical mutational dynamics in some runs and are due to the combined effects of using a constant set of 20 seeds for all runs and the relatively weak effect of selection in small populations. As in Figure 4, $s_M = 0.05$, $s_A = 0.05$, and the mutation rates are 10^{-6}.

described above, in which the $A_1A_1B_2B_2$ or $A_2A_2B_1B_1$ genotype fixes before adaptive evolution begins, no more than 2/3 utilization of the direct path is expected. Indeed, cohorts of small ($N = 100$) and medium ($N = 1000$) populations utilize both pathways (data not shown). The rapidity of DM3 in large populations is therefore a combined effect of the fitness model and polymorphism, which together promote use of the fastest route to pseudocompensation.

Simulations using DM2, the BPD model in which s_A is awarded solely for gain of bond 3, also generally support the predictions of the analysis (Fig. 5, center left panel). For example, in the largest populations the initial, maximal rate of pseudocompensation is about half that seen in DM1. But once the eight of 20 populations taking the direct route reach semi-incompatibility, the remaining populations, which use the indirect path, evolve markedly slower. DM2 also shows clear population size dependence, but to a lesser extent than DM1 or DM3. This is presumably because, although DM2

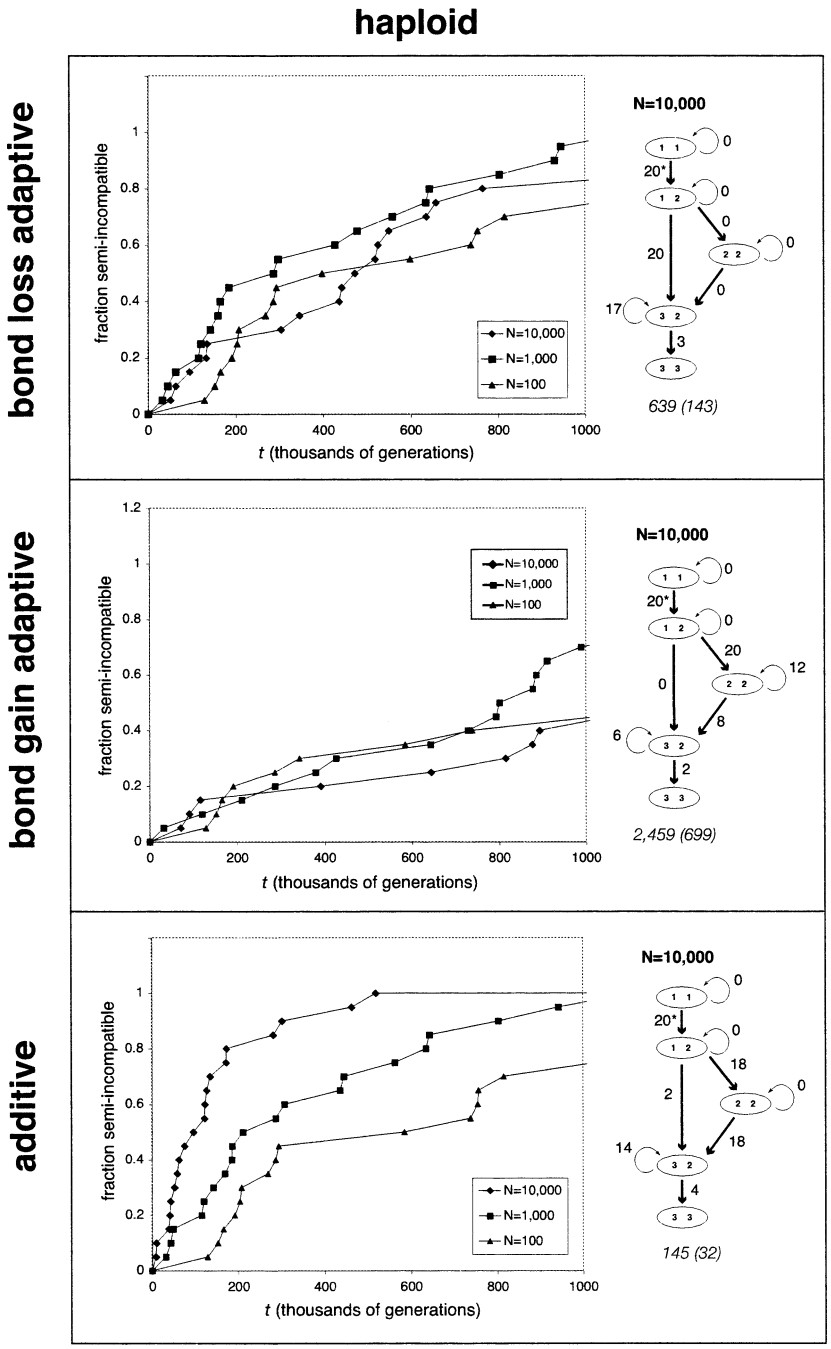


FIG. 6. Haploid adaptive pseudocompensation. Populations of 100, 1000, and 10,000 haploid individuals were tracked as they evolved to the semi-compensated genotype (A_2B_3 or A_3B_2) as in Figure 5. To the right of each accumulation plot is a summary of the allelic substitutions that occurred during the first million generations of evolution in the $N = 10{,}000$ runs. The three scenarios for adaptive benefit of the semi- or fully compensated genotype are noted on the left. The asterisks after the number of populations taking the first mutational step indicate that in most or all of these simulations the A_1B_2 or A_2B_1 genotype failed to reach fixation before the A_2B_2 genotype began to increase. As in Figures 5 and 6, $s_M = 0.05$, $s_A = 0.05$, and the mutation rates are 10^{-6}.

is a BPD model, the heavy use of the indirect pathway with its mutation rate-dependent rate-limiting step dampens population size effects. Interestingly, DM2 is actually significantly slower (*t*-test, $P = 0.03$) to produce incompatibility in large populations than is the two-allele BPD model (compare Fig. 5, center-left panel, with Fig. 4A). We note that the direct path for compensatory change used in the two-allele BPD model is still open to populations in the three-allele simulations. That DM2 produces incompatibility more slowly than the two-allele model demonstrates that, contrary to expectation, the intermediates in some cases actually interfere with the truly compensatory evolution of incompatibilities.

As with the two-allele models, codominance impedes the

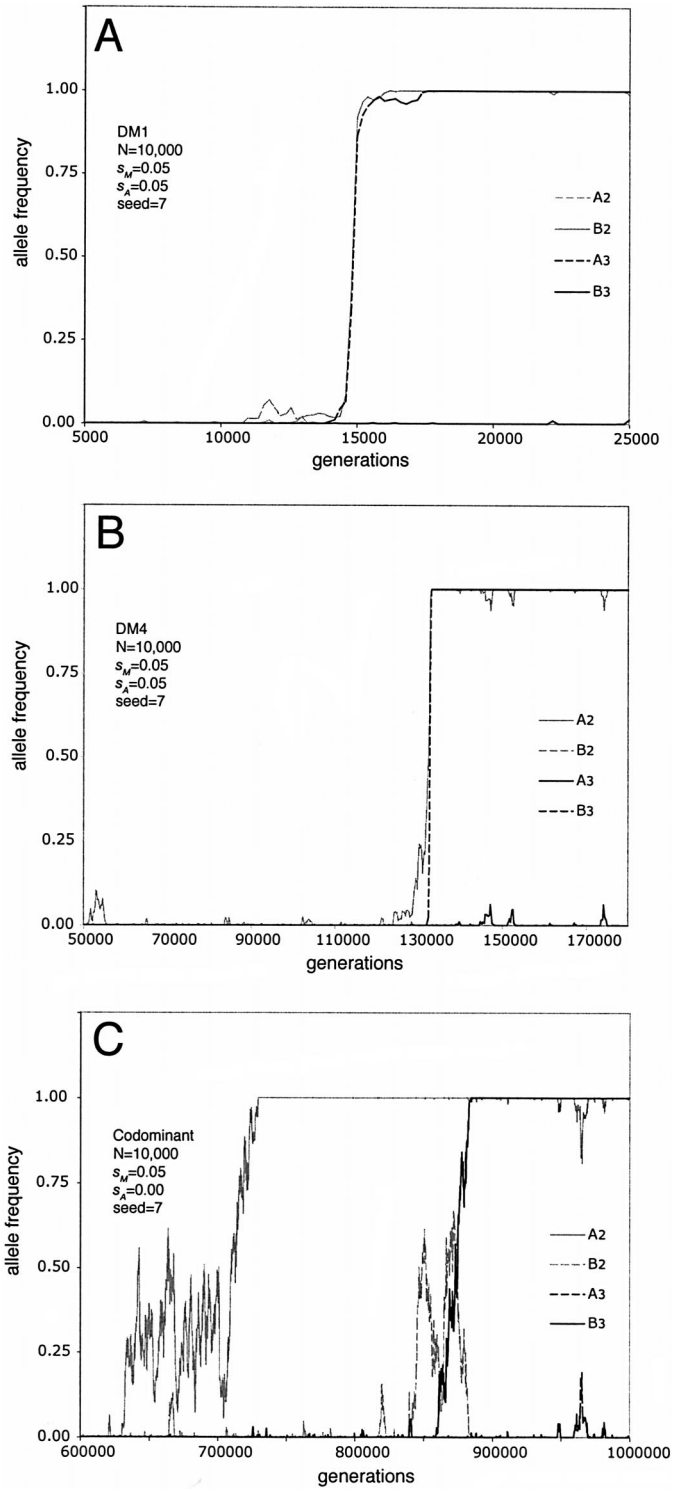


FIG. 7. Examples of allele frequency dynamics in the best pair dominant (BPD) and codominant models. (A) Under DM1, a BPD model which favors loss of bond 1, 3-type alleles can rapidly spread when even a small population of the other locus' 2-type allele is present. In this case, a very small proportion of allele B_2 has been produced by drift at around generation 14,000. The subsequent arising of the A_3 mutation in this context leads to rapid fixation of both alleles. (B) Under DM4, a codominant model that also favors loss of bond 1, the intermediate allele (in this case A_2) reaches a relatively high frequency by drift (roughly 25% at generation 13,000) before the compensatory mutation (B_3) is able to gain a foothold. (C) Simulation using the same conditions as in (A) and (B) except without selection for the pseudocompensated genotype (i.e., $s_A = 0$). Fixation of A_2 and B_3 are now uncoupled greatly in time, and semi-incompatibility is reached much later overall. Note that the scale of generations (x-axis) and the line style for each allele are different in each graph.

evolution of incompatibility relative to the BPD scenario. Why this is the case is revealed by examining the allele frequencies of individual large-population runs, as shown in Figure 7 (A, B). In this representative example, the speed of DM1 (BPD) pseudocompensation is about nine-fold faster than DM4 (codominant). As noted above, this is because in the BPD model DM1 the initially rare A_3-B_2 genotype gets a selective advantage, yet the A_3 allele is not selected against when placed in the more common A_1 background by recombination. But in DM4, the second, 3-type mutation (B_3 in this case) can only survive when it coincides with a drift-based spike in A_2 frequency, which is necessary to overcome negative selection when paired with A_1. Thus, the waiting here is not just for the compensatory mutation, but also for the appropriate neutral mutation to become abundant enough to make its survival likely.

The worst diploid model overall for pseudocompensation is DM5, the codominant scenario in which bond gain is adaptive (Fig. 5, center right panel). This model is expected to be slow because it lacks the features that made DM1 and DM3 work well. That is, codominance prevents A_3 or B_3 alleles from increasing when rare unless a sizeable fraction of B_2 or A_2 alleles (respectively) are already present, and the adaptive bond gain forces increased use of the indirect mutational path. Surprisingly, however, the mean time to semi-incompatibility under DM5 is actually negatively correlated with population size (data not shown), the only model for which this is the case. Population size is also related to the tendency to take the indirect mutational path: in the simulations with large populations (N = 10,000), 100% of populations use it (Fig. 5, center-right panel), as opposed to the roughly 50% indirect path usage seen in the small and medium population size DM5 cohorts (not shown). As the indirect path is relatively slow, its exclusive use underlies the overall slowness of pseudocompensation.

The correlation between population size and use of the indirect path in the DM5 model can be explained as follows: With increasing population size, A_2 or B_2 alleles more frequently pair with the second, synthetically adaptive allele while still polymorphic and relatively rare. With codominance, this adaptive allele pair will spread more easily if neither allele is deleterious in combination with the more abundant ancestral alleles. Given their mutual rarity and lack of linkage, this will favor fixation of the $A_2A_2B_2B_2$ double intermediate over $A_2A_2B_3B_3$ or $A_3A_3B_2B_2$, even though all three have the same fitness. Having thus been forced into the indirect path, the entire process is slowed. However, we note that even in this worst case model, it is still much more effective at allowing the evolution of incompatibility than is a two-allele codominant model (see Fig. 4).

Of the three codominant models, the best at promoting pseudocompensation is DM6, in which adaptation is an ad-

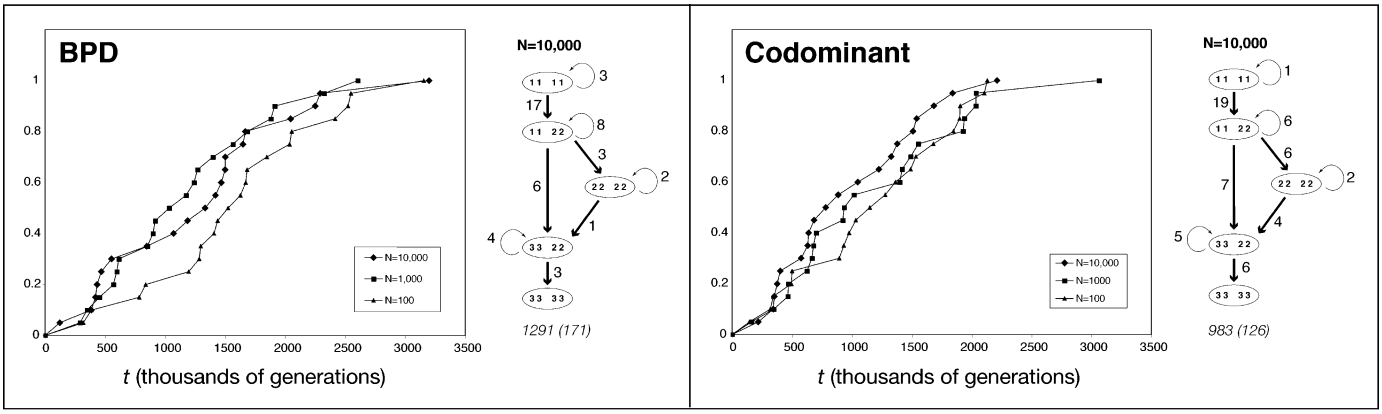


FIG. 8. Neutral pseudocompensatory evolution with intermediate alleles. Diagrams are labeled as described in Figures 5 and 6, but note that the range of times on the x-axis is 3.5 times greater. Without adaptation ($s_A = 0$) incompatibility still evolves, even in the codominant case, but it does so much more slowly than in the adaptive three-allele case (Fig. 5). The allele substitution diagrams, on the right of each panel, show that even large populations ($N = 10{,}000$) are often fixed at any of the intervening genotypes at the end of 1 million generations. They also show that the direct and indirect paths are used approximately equally.

ditive effect of bond loss and gain (Fig. 5, lower right panel). In large populations, DM6 enables incompatibility to evolve at a rate similar to the two-allele BPD model, and about threefold slower than the diploid BPD models DM1 and DM3. Given that 75% of the populations in the large-size cohort followed the indirect mutational path, at first glance unclear why it should be superior to DM4, the adaptive bond-loss scenario, in which 100% take the direct route. However, scrutiny of the details of each simulation shows that the waiting required for a successful direct path transition under DM4 (described above and in Fig. 7) is avoided in DM6. In each case where the indirect path is used under DM6, the $A_2A_2B_2B_2$ intermediate reaches fixation via a synergistic and simultaneous increase in the frequency of the A_2 and B_2 alleles. This strong effect of polymorphism is more reliable in large populations. However, unlike DM5, in which populations also heavily utilize the indirect path, the additive model employed by DM6 allows an adaptive step from the $A_2A_2B_2B_2$ double intermediate to either semi-incompatible genotype (see Fig. 3A). As a result, the indirect path is greatly speeded in a population size-dependent fashion.

To further clarify the role of dominance in the diploid simulations, we also performed haploid simulations with the three adaptive fitness models (Fig. 6). The results closely parallel those of the equivalent codominant diploid simulations, but show overall slower evolution of semi-incompatible genotypes. This general trend is expected, as the total number of alleles, and thus the rate of adaptive evolution, is halved in haploids. Comparison of the simulation data from the diploid and haploid versions of each fitness model supports this. For example, in the adaptive bond gain scenario, codominant diploid populations fix the $A_2A_2B_2B_2$ genotype in approximately half the time it takes haploid populations to fix at A_2B_2.

For the haploid implementation of adaptive bond loss and adaptive bond gain scenarios, evolution is essentially independent of population size. In the first case, this is due to the minimal role for polymorphism in accelerating adaptive evolution. Only the direct path is adaptive, and this must occur by sequential fixation (or near-fixation) of an intermediate allele before the adaptive 3-type allele can begin increasing. In the second case, this is because all medium and large populations take the indirect path, and must wait long periods to drift from A_2B_2 to either A_2B_3 or A_3B_2 (similar to DM5). The haploid simulations of the additive model show the same accelerated evolution in larger populations that were seen with DM6 (the codominant equivalent). Again, this is due to the ability of both A_2 and B_2 to reinforce each other while still relatively rare without any cost of mismatch in the B_1 or A_1 background, and to the lack of a neutral step in the indirect path. The additive model thus behaves as well or better than the adaptive bond gain or loss scenarios across the board when population sizes are large.

Neutral Pseudocompensatory Evolution with Intermediates

To understand the magnitude of the role of positive selection in the adaptive runs, we performed dipoid simulations in which $s_A = 0$ (Fig. 8). These neutral runs reveal that the existence of intermediate alleles alone can facilitate pseudocompensatory evolution, but also underscore the importance of selection in speeding the process. As with the adaptive runs, the inclusion of intermediates allowed the evolution of ancestor-incompatible genotypes, even in small populations. Unlike the adaptive runs, however, no more than 55% of runs for any given set of conditions reached the semi-incompatible state in the first 1 million generations, and the pseudocompensatory evolution that does occur is very slow.

As shown in Figure 7C, the time that alleles are polymorphic is greatly increased in the neutral case. As expected, neither the mean number of generations required to reach incompatibility with the ancestor nor the number of individual runs that get there in 1 million generations is correlated with population size. There is also no clear effect of dominance in neutral pseudocompensation, as judged by the similar accumulation of semi-incompatible populations over time in BPD and codominant simulations (Fig. 8). The long time scale involved in neutral pseudocompensation generally imposes strict sequential fixation of alleles. As a result, the

benefit of dominance in shielding mismatches alleles is rendered moot.

Rerunning the neutral simulations for 5 million generations, long enough for all 20 replicates to reach at least a semi-compensated genotype, enabled rough estimation of the enhancement of pseudocompensatory change by a particular selection coefficient for a given population size and fitness model. The enhancement for the large selection coefficients we studied ($s_M = 0.05$, $s_A = 0.05$) ranges from less than 30% for DM5 to 41-fold for DM1, with both extremes coming from the largest ($N = 10{,}000$) populations.

Discussion

We have explored the effects of introducing intermediate adaptor alleles on the evolution of incompatibilities in interacting gene products. If intermediate alleles are effective, they should facilitate such evolution where it is otherwise essentially impossible, and speed it up under conditions where it would otherwise be slow. Our results show that they can accomplish both of these things, even as they necessitate more mutational steps to reach incompatibility. In particular, intermediates allow pseudocompensatory evolution in small populations, in the absence of adaptation, in haploid populations, and in diploid populations under codominant fitness models. The intermediates also greatly accelerate the process in the BPD diploid models. Furthermore, the intermediates provide a path by which unlinked loci can reliably coevolve.

Crow and Kimura (1965) modeled haploid sexual populations undergoing traditional two locus/two allele compensatory evolution. They note that free recombination requires that the selective benefit of the derived allele pair be strong enough to withstand negative selection when either member of the pair (i.e., A_3 or B_3 in the language of this paper) is placed in a more common A_1/B_1 background. Specifically, s_A must be greater than $r/(1 - r)$, where r is the fractional recombination between loci each generation, for the A_3B_3 genotype to fix deterministically. Thus, with free recombination (i.e., $r = 0.5$) the compensated genotype must be more than twice as fit as the ancestral state, presumably a tall order. We have demonstrated here that the process of pseudocompensation allows for predictable fixation when $r = 0.5$ but s_A is only 0.05, 20-fold weaker than the selection coefficient required under the above condition. Although a systematic exploration of the interaction of recombination and positive selection in pseudocompensation is beyond the scope of this paper, it is clear that lower values of s_A could also allow the evolution of incompatibility.

Gene Product Structure, Adaptation, and Polymorphism in Compensatory Evolution

Compared with the two-allele models, the intermediate alleles generally promote the evolution of adaptive, ancestor-incompatible genotypes. However, in large populations, the specific aspect of an interaction that is adaptive has a large effect. In most cases intermediate alleles accelerate the evolution of incompatibility, but there is one interesting exception. Under the adaptive bond gain scenario and BPD diploidy (DM2), intermediates actually slow the evolution of incompatibility relative to the two-allele case by nearly three-fold. Adaptive benefit dependent upon gain of a new bond, as in DM2, is granted regardless of whether a redundant bond is lost or not. The consequence of such selection is the increased persistence of the multifunctional intermediate alleles, and a decrease in specificity. In contrast, benefit dependent upon loss of a preexisting bond would promote specificity, and might be especially important where reduction of pleiotropic interactions or an increase in molecular crypsis is important.

Our model predicts that polymorphisms in or near the interacting surfaces of dimeric proteins are key facilitators of compensatory (or pseudocompensatory) evolution. To be effective, these polymorphisms should constitute all or part of unused half-bonds, which are alone neutral or nearly neutral. Unlike an adaptive compensatory change in the binding partner, which should rapidly reach fixation once common, these intermediate polymorphisms may be present for long periods of time. The model also indicates that after semi-incompatibility evolves, there is also a long delay (typically hundreds of thousands of generations) between fixation of $A_2A_2B_3B_3$ or $A_3A_3B_2B_2$ and fixation of $A_3A_3B_3B_3$ by drift in all fitness models. This suggests that the mutations that finalize incompatibility can also be polymorphic for long periods. This may be relevant to recent studies of *Tribolium* anatomy and *Caenorhabditis* sex determination that show intraspecies variation in the propensity to form viable interspecies hybrids (Wade et al. 1997; Baird 2002). The latter case may involve incompatibilities in known interactions between sex determination gene products (Chin-Sang and Spence 1996; Mehra et al. 1999; Wang and Kimble 2001; Haag et al. 2002)

Excess Capacity as a Facilitator of Evolution

Our model depends heavily upon the possibility that gene products can have the potential to interact with more or different partners than is required at any given time for performance of their key functions. This sort of excess capacity can then be shed in a number of ways, some of which lead to ancestor-descendant incompatibility. In this way, our model is similar to those of gene duplication and divergence (Force et al. 1999; Stoltzfus 1999; Lynch and Force 2000). It is also similar to studies of the evolution of gene regulatory circuits under stabilizing and directional selection elaborated by Johnson and Porter (2000; Porter and Johnson 2002). In all of these cases, the two keys to evolution are (1) the multiplicity of ways in which an important molecular function can be achieved; and (2) little or no negative fitness effects of potential intermediate alleles harboring excess capacity.

Both of the above requirements seem reasonable in light of available evidence. The rate of gene duplication in eukaryotic genomes has been estimated, using a demographic approach, to be roughly 10^{-7} per gene per year, similar to the per nucleotide mutation rate (Lynch and Conery 2000). For interacting proteins, the great potential for intermediate alleles is supported by both the numerous cases of rapidly coevolving proteins and nucleic acids described in the introduction, and also by the high levels of amino acid polymorphism found in some interacting proteins (e.g., Haag and Ackerman 2005). However, in other cases selective sweeps may remove all evidence of transient intermediates (Metz et al. 1998).

Evolutionary Structural Biology

To test the relevance of this model of compensatory evolution, a new kind of comparative structural biology is required. Current priorities for crystallography studies emphasize the discovery of unique structural folds. However, to understand compensatory evolution, the determination of structures of closely related interacting proteins is required, ideally as heterodimers. Such structures can further be used to map amino acid polymorphisms within a species that may represent a nascent or leftover intermediate. One recent example of this kind of work is the model-based structural studies of polymorphisms in one of the two interacting proteins comprising the rapidly coevolving *Brassica* self-incompatibility system (Chookajorn et al. 2004). With more such data, the details of how protein-protein interactions are maintained in spite of sequence change will become apparent.

Acknowledgments

We thank S. Hsieh, A. Stoltzfus, and N. Johnson for useful discussions and A. Doty and two anonymous reviewers for their helpful comments on the manuscript. We are grateful to K. Zhang and the University of Maryland Institute for Advanced Computer Studies for invaluable assistance with performing the simulations. This work was supported by start-up funds from the University of Maryland to ESH.

Literature Cited

Baird, S. 2002. Haldane's rule by sexual transformation in *Caenorhabditis*. Genetics 161:1349–1353.

Barton, N. 1989. The divergence of a polygenic system subject to stabilizing selection, mutation, and drift. Genet. Res. Camb. 54: 59–77.

Carter, A., and G. Wagner. 2002. Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc. R. Soc. Lond. B 269:953–960.

Chin-Sang, I. D., and A. M. Spence. 1996. *Caenorhabditis elegans* sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. Genes Dev. 10:2314–2325.

Chookajorn, T., A. Kachroo, D. Ripoli, A. Clark, and J. Nasrallah. 2004. Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. Proc. Natl. Acad. Sci. USA 101:911–917.

Crow, J., and M. Kimura. 1965. Evolution in sexual and asexual populations. Am. Nat. 99:439–450.

Force, A., M. Lynch, F. Pickett, A. Amores, Y. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

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

Gillespie, J. 1984. Molecular evolution over the mutational landscape. Evolution 38:1116–1129.

Haag, E., and A. Ackerman. 2005. Intraspecific variation in *fem-3* and *tra-2*, two rapidly coevolving nematode sex-determining genes. Gene 349:35–42.

Haag, E., S. Wang, and J. Kimble. 2002. Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. Curr. Biol. 12: 2035–2041.

Haldane, J. 1931. A mathematical theory of natural selection. VIII. Metastable populations. Proc. Camb. Philos. Soc. 27:137–143.

———. 1932. The causes of evolution. Cornell Univ. Press, Ithaca, NY.

Hellberg, M., and V. Vacquier. 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. Mol. Biol. Evol. 16:839–848.

Iwasa, Y., F. Michor, and M. Nowak. 2004. Stochastic tunnels in evolutionary dynamics. Genetics 166:1571–1579.

Johnson, N., and A. Porter. 2000. Speciation via parallel, directional selection on regulatory genetic pathways. J. Theor. Biol. 205: 527–1542.

Kachroo, A., C. Schopfer, M. Nasrallah, and J. Nasrallah. 2001. Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293:1824–1826.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, U.K.

———. 1985. The role of compensatory neutral mutations in molecular evolution. J. Genet. 64:7–19.

———. 1990. Some models of neutral evolution, compensatory evolution, and the shifting balance process. Theor. Popul. Biol. 37:150–158.

Kondrashov, A., S. Sunyaev, and F. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. USA 99:14878–14883.

Lynch, M., and J. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155.

Lynch, M., and A. Force. 2000. The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. 156:590–605.

Mehra, A., J. Gaudet, L. Heck, P. E. Kuwabara, and A. M. Spence. 1999. Negative regulation of male development in *Caenorhabditis elegans* by a protein-protein interaction between TRA-2A and FEM-3. Genes Dev. 13:1453–1463.

Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol. Biol. Evol. 13:397–406.

Metz, E. C., R. Robles-Sikisaka, and V. D. Vacquier. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceed substitution in introns and mitochondrial DNA. Proc. Natl. Acad. Sci. USA 95:10676–10681.

Michalakis, Y., and M. Slatkin. 1996. Interaction of selection and recombination in the fixation of negative-epistatic genes. Genet. Res. Camb. 67:257–269.

Phillips, P. 1996. Waiting for a compensatory mutation: phase zero of the shifting-balance process. Genet. Res. Camb. 67:271–283.

Porter, A., and N. Johnson. 2002. Speciation despite gene flow when developmental pathways evolve. Evolution 56:2103–2111.

Rousset, F., M. Pelandakis, and M. Solignac. 1991. Evolution of compensatory substitutions through G:U intermediate stat in *Drosophila* rRNA. Proc. Natl. Acad. Sci. USA 88:10032–10036.

Stephan, W. 1996. The rate of compensatory evolution. Genetics 144:419–426.

Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. J. Mol. Evol. 49:169–181.

Swanson, W., and V. Vacquier. 1995. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc. Natl. Acad. Sci. USA 92:4957–4961.

———. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137–144.

Wade, M., N. Johnson, R. Jones, V. Siguel, and M. McNaughton. 1997. Genetic variation segregating in natural populations of *Tribolium castaneum* affecting traits observed in hybrids with *T. freemani*. Genetics 147:1235–1247.

Wang, S., and J. Kimble. 2001. The TRA-1 transcription factor binds TRA-2 to regulate sexual fates in *Caenorhabditis elegans*. EMBO J. 20:1363–1372.

Wilson, E., and W. Bossert. 1971. A primer of population biology. Sinauer Associates, Sunderland, MA.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

———. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. 6th Int. Congr. Genet. 1: 356–366.

Corresponding Editor: P. Phillips