

RAPID EVOLUTIONARY ESCAPE BY LARGE POPULATIONS FROM LOCAL FITNESS PEAKS IS LIKELY IN NATURE

DANIEL M. WEINREICH^{1,2} AND LIN CHAO^{3,4}

¹Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138

²E-mail: dmw@post.harvard.edu

³Division of Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093

⁴E-mail: lchao@biomail.ucsd.edu

Abstract.—Fitness interactions between loci in the genome, or epistasis, can result in mutations that are individually deleterious but jointly beneficial. Such epistasis gives rise to multiple peaks on the genotypic fitness landscape. The problem of evolutionary escape from such local peaks has been a central problem of evolutionary genetics for at least 75 years. Much attention has focused on models of small populations, in which the sequential fixation of valley genotypes carrying individually deleterious mutations operates most quickly owing to genetic drift. However, valley genotypes can also be subject to mutation while transiently segregating, giving rise to copies of the high fitness escape genotype carrying the jointly beneficial mutations. In the absence of genetic recombination, these mutations may then fix simultaneously. The time for this process declines sharply with increasing population size, and it eventually comes to dominate evolutionary behavior. Here we develop an analytic expression for N_{crit} , the critical population size that defines the boundary between these regimes, which shows that both are likely to operate in nature. Frequent recombination may disrupt high-fitness escape genotypes produced in populations larger than N_{crit} before they reach fixation, defining a third regime whose rate again slows with increasing population size. We develop a novel expression for this critical recombination rate, which shows that in large populations the simultaneous fixation of mutations that are beneficial only jointly is unlikely to be disrupted by genetic recombination if their map distance is on the order of the size of single genes. Thus, counterintuitively, mass selection alone offers a biologically realistic resolution to the problem of evolutionary escape from local fitness peaks in natural populations.

Key words.—Compensatory mutation, epistasis, fitness landscape, genetic recombination, mean fitness, shifting balance theory.

Received June 24, 2004. Accepted March 10, 2005.

Because functional interactions between mutations are common, the fitness effect of a mutation occurring at one locus may depend on the allelic state at one or more functionally interacting loci, a circumstance called epistasis for fitness. Sewall Wright (1932) noted that as a consequence of epistasis, groups of mutations might exist that when introduced jointly into a genotype increase fitness but which reduce fitness if introduced individually. On the discrete genotypic fitness landscape (Wright 1932) such interactions are manifest as a local peak, all of whose immediate mutational neighbors in genotypic sequence space (analogous to the “protein space” of Maynard Smith 1970) have lower fitness in spite of the existence of a higher fitness genotype two or more mutations removed (Fig. 1 illustrates the biallelic two locus case). Wright (1932, p. 358) felt that there would be an “enormous number of widely separated harmonious [allelic] combinations . . . separated by [low fitness] valleys” and empirical examples of such fitness interactions have been demonstrated among mutations in bacteria (Schrag et al. 1997; Maisnier-Patin et al. 2002) and in viruses (O’Hara et al. 1984; Giachetti and Holland 1988; Burch and Chao 2000). Moreover, it should be noted that the absence of similarly explicit examples from multicellular organisms may only reflect methodological limitations in the construction and fitness assessment of mutationally adjacent genotypes and inferential evidence for multiple fitness peaks in metazoans exists (e.g., Whitlock et al. 1995; Kondrashov et al. 2002).

The problem of how populations might fix two or more mutations that are individually deleterious but jointly beneficial has been a puzzle in evolutionary genetics for 75 years (Wright 1930, 1932; Provine 1986). Presumably following

the common intuition that the evolutionary escape from such local peaks would require the sequential fixation of a succession of low-fitness valley genotypes before the higher-fitness escape genotype could be produced by mutation, Wright focused attention on models of population subdivision (Wright 1930, 1932). Here natural selection against valley genotypes within demes is weakened by genetic drift, thereby easing the transit, and much subsequent attention has focused on this stochastic sequential process (e.g., Wright 1965, 1970; Crow 1991; Wade and Goodnight 1991; Moore and Tonsor 1994).

However, no mutationally intermediate valley genotypes need fix before the high-fitness escape genotype appears in an evolving population because unless lethal, valley genotypes are expected to persist for a few generations before they are selectively eliminated, and during this time interval they may be subject to mutation at other loci. Thus, before any valley genotypes fix, the escape genotype carrying multiple mutations that are beneficial only jointly may arise, and in the absence of genetic recombination this genotype may then be swept to fixation by natural selection. The possibility of this alternate population genetic process in which the jointly beneficial mutations fix simultaneously was first suggested by Gillespie (1984). He noted that the rate of production of escape genotypes would be proportional to population size, and because the probability of fixation of a beneficial allele is largely independent of population size (Haldane 1927), Gillespie (1984) recognized that under this process the expected time for evolutionary escape would decline with increasing population size. This deterministic behavior stands in sharp contrast to that for the stochastic sequential fixation

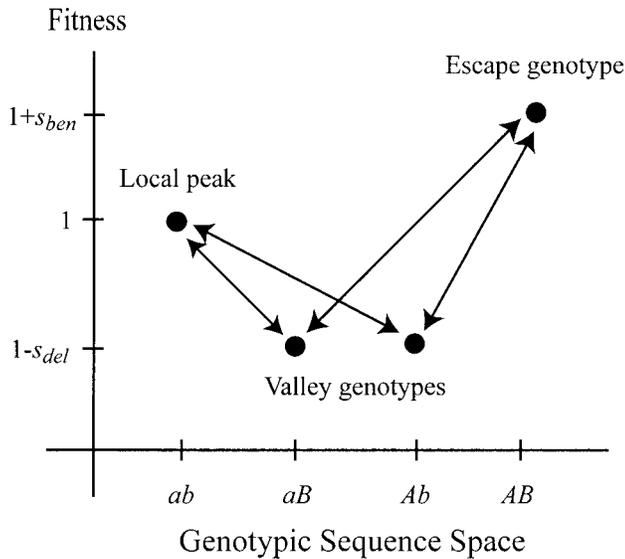


FIG. 1. Fitness model employed. Genotypes defined for two biallelic loci, arranged along the x-axis. Fitness of each genotype is represented by filled circles, with fitness values shown on y-axis. Arrows illustrate mutational pathways. Genotype ab is a local peak; aB and Ab represent valley genotypes and AB represents the escape genotype.

process in which the expected time for escape increases with population size.

Carter and Wagner (2002) showed a distinctly nonmonotonic response to population size in the rate of escape from a local peak in a diploid nonrecombining model incorporating both these processes. In sufficiently small populations the rate declines as size increases, as expected for the drift-dependent stochastic sequential fixation process, but eventually the rate of escape increases dramatically as the deterministic simultaneous fixation process comes to dominate population evolution. Very similar dynamics were reported by Iwasa et al. (2004), who employed a branching process model to find an expression for the probability of transit across a fitness valley as a function of time (these authors denote the simultaneous fixation process “stochastic tunneling”).

In addition to population size, evolutionary dynamics across a fitness valley are sensitive to recombination rate (Crow and Kimura 1965; Phillips 1996). In small populations in which mutations fix sequentially, recombination is unlikely to matter because two loci will rarely be simultaneously polymorphic (Stephan 1996). However, in populations large enough to permit the simultaneous fixation of jointly beneficial mutations, frequent recombination may destroy escape genotypes before they reach fixation, in which case escape from a local peak on the genotypic fitness landscape will again be stochastic (Phillips 1996). This stochastic process differs from the sequential fixation process operating in small populations because the fixation of low-fitness valley genotypes is not required. Instead, valley genotypes need only drift to sufficient frequency that the production of escape genotypes via recombination more than offsets their destruction by recombination, after which the jointly beneficial mutations again fix simultaneously. Nevertheless, under this stochastic simultaneous fixation process the expected time for

evolutionary escape from a local peak on the genotypic fitness landscape goes up with increasing population size, owing again to the dependence on genetic drift.

Thus, sensitivity in escape dynamics exists both to population size and to recombination rate, and formally three regimes appear possible. In sufficiently small populations evolutionary escape is via the stochastic sequential fixation of one or more low-fitness valley genotypes followed by the appearance and fixation of the escape genotype; in large populations these jointly beneficial mutations will fix simultaneously. However, whereas in the absence of recombination this process is deterministic, for sufficiently large recombination rate the process will again be stochastic. Which of these regimes are likely to exist in natural populations?

We first present a general and intuitively simple analytic model for the expected time to escape from a local peak on the genotypic fitness landscape as a function of population size assuming complete genetic linkage. This treatment is robust to depth and width of fitness valley, fitness gain across the valley and mutation rate, and is well corroborated by individual-based computer simulation in which model assumptions are relaxed. We next develop an expression for the critical population size that defines the boundary between the sequential and simultaneous fixation regimes, and for biologically plausible parameter values this threshold may be quite small. Finally, simulation results motivate a novel expression for the critical recombination rate above which the deterministic simultaneous fixation of jointly beneficial mutations gives way to their stochastic simultaneous fixation. Together with recent estimates of the per nucleotide recombination rate in metazoans, this threshold value suggests that at least at the scale of individual genes, epistatically interacting mutations such as shown in Figure 1 may not be disrupted frequently enough to undermine their deterministic fixation. In summary, we conclude that the deterministic simultaneous fixation of mutations only jointly beneficial is likely in natural populations, and thus counterintuitively that evolutionary escape from a local peak on the genotypic fitness landscape may be fastest in large populations.

THE MODEL

By definition, a local fitness peak is a genotype with no individually beneficial mutations available in spite of one or more genotypes of higher fitness two or more mutations removed on the genotypic fitness landscape. The simplest possible case consists of two biallelic loci that define a fitness valley two mutations wide lying between two higher fitness genotypes, as shown in Figure 1. Let the ab genotype occupy a local peak with fitness 1. The mutationally intermediate valley genotypes aB and Ab occupy the fitness valley and have fitness $1 - s_{del}$ ($0 < s_{del} < 1$). Finally, let the escape genotype AB have the highest fitness, $1 + s_{ben}$ ($s_{ben} > 0$).

We model a panmictic population of N diploid individuals with genic selection or equivalently a haploid population of $2N$ individuals. Generations are discrete and mutations occur at each locus with probability μ per chromosome per generation. Following Gillespie (1984), we make the strong selection/weak mutation assumption that $s_{del}, s_{ben} \gg 1/N \gg \mu$, so the time to fixation of a mutation will be dominated by

the time for the mutations to appear and escape genetic drift. Thus, we disregard the transit time of a mutation through frequency space on its way to fixation and approximate the expected time to fixation of a mutation by the reciprocal of the per generation probability of the appearance of a mutation destined to fix (Gillespie 1984; Stephan 1996).

MONTE CARLO COMPUTER SIMULATIONS

To test the assumptions underlying our analytic results, individual-based Monte Carlo computer simulations (available upon request) were performed. Populations were founded with N diploid individuals homozygous for ab chromosomes. In each generation a Poisson-distributed number ($\lambda = 2N\mu$) of mutations at each locus were assigned to chromosomes sampled with replacement, following the mutational pathways illustrated in Figure 1. Chromosome counts in the next generation were then determined stochastically via multinomial sampling (Gillespie 1993) using chromosome frequencies adjusted deterministically to reflect recombination between loci (Ewens 1979) and genic selection (Crow and Kimura 1970). (The stochastic introduction of recombination events did not affect results, not shown.) This cycle was iterated until the entire population was homozygous for the AB chromosome, and the number of generations elapsed was recorded. We denote the mean number of generations elapsed across replicate evolutionary Monte Carlo realizations as T_{MC} ; results are for 1000 replicates except as noted.

RESULTS

Beginning with a population fixed for ab chromosomes, we seek the expected time in generations until the population fixes the higher fitness AB chromosome. Two population genetic processes are possible (Carter and Wagner 2002). In the stochastic sequential case, a population fixes one of the mutationally intermediate chromosomes (aB or Ab) and subsequently fixes the AB chromosome. The expected time for this process is

$$T_{seq} = \frac{1}{4N\mu \cdot u(N, -s_{del})} + \frac{1}{2N\mu \cdot u(N, s_{del} + s_{ben})}, \quad (1)$$

where $u(N, s)$ is the probability of fixation of a mutation with genic selection coefficient s in a diploid population with effective size N , given by $(1 - e^{-2s})/(1 - e^{-4Ns})$ (Kimura 1962). The first term in equation (1) represents the time to fix either the aB or Ab chromosome, equal to the reciprocal of the product of the rate of mutation (at either of two loci in either of two chromosomes in each individual) and the probability of fixation. The second term represents the time to fix the AB chromosome in a population fixed for either aB or Ab . Here the rate of production is halved because only a single locus remains in the unmutated state (we disregard back mutation in eq. 1), and the probability of fixation reflects the larger fitness gain between aB or Ab and AB .

Alternatively, under the deterministic simultaneous fixation process, the AB chromosome appears via mutation of a segregating mutationally intermediate valley chromosome. Disregarding recombination between loci, the expected time for this process is given by

$$T_{sim} = \frac{s_{del}}{4N\mu^2 \cdot u(N, s_{ben})}. \quad (2)$$

Here, the equilibrium frequency of valley chromosomes is assumed to be $2\mu/s_{del}$, the mutation-selection balance (Kimura 1985b; Stephan 1996) under genic selection, again accounting for two loci. Only the fraction of mutations that occur in this subset of chromosomes, of which there are $\sim 4N\mu/s_{del}$ copies, will create AB chromosomes. Thus, equation (2) may be seen as the reciprocal of the product of the rate of production of AB chromosomes by the mutation of segregating valley chromosomes and their probability of fixation.

We can explicitly disregard the possibility of fixation of an AB chromosome generated by simultaneous mutation at both loci in a single generation. Following the same strong selection/weak mutation approach, the expected time for escape via simultaneous mutation at both loci is

$$\frac{1}{2N\mu^2 \cdot u(N, s_{ben})},$$

which exceeds equation (2) by the factor $2/s_{del} > 1$. A simple explanation for this relationship is possible: by the definition of s_{del} , each valley chromosome persists on average for $1/s_{del}$ generations before elimination by purifying selection, and two valley chromosomes are defined. Thus, to give rise to an AB chromosome via the process modeled in equation (2), the second mutation may occur anytime during a window effectively $2/s_{del}$ generations long; to produce the AB chromosome via simultaneous mutation the analogous window is only one generation long.

Finally, we denote as T_{esc} the expected time for a population fixed for the ab chromosome to escape from the local peak and fix the higher fitness AB chromosome via either process. This is given by

$$T_{esc} = \frac{1}{1/T_{seq} + 1/T_{sim}} \quad (3)$$

(Carter and Wagner 2002) and reflects the fact that the expected time for any process is the reciprocal of its rate (Stephan 1996). The net rate at which a population will escape via either process is the sum of the individual rates, the reciprocal of which is thus the expected net time.

T_{seq} , T_{sim} and T_{esc} (dotted, dashed, and solid lines, respectively) are shown as functions of N in Figure 2 for $s_{del} = 10^{-2}$, $s_{ben} = 10^{-1}$, and $\mu = 10^{-5}$. T_{seq} grows almost monotonically in N because as population size increases the probability of fixation of either low-fitness valley chromosome declines. Conversely, T_{sim} declines in N because the number of segregating low-fitness valley chromosomes increases. (Unlike T_{seq} , which is reduced for small N owing to weakened selection, T_{sim} does not decline for large N because of weakened drift, since as N increases, $u(N, s)$ quickly approaches $2s$ for $s > 0$.) These functions cross at some N_{crit} . Because T_{esc} is dominated by the smaller of its two constituent terms, $T_{esc} \approx T_{seq}$ for $N < N_{crit}$, while $T_{esc} \approx T_{sim}$ for $N > N_{crit}$. Thus, N_{crit} defines the boundary between the sequential and simultaneous fixation regimes.

The form of results shown in Figure 2 is robust to parameter values. For

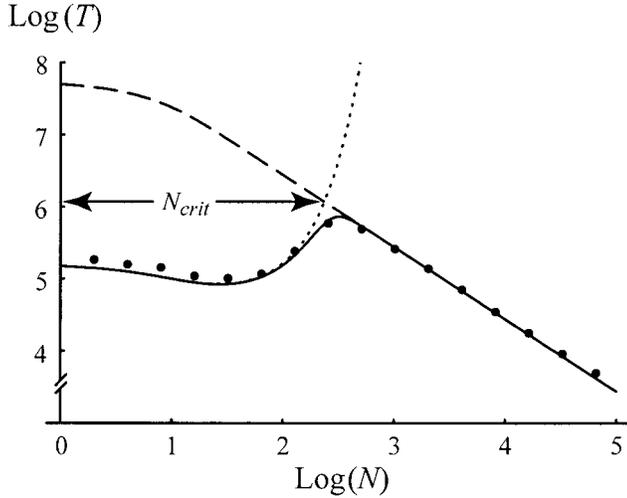


FIG. 2. The expected time T for populations to escape a low fitness peak as a function of population size N . Expectation under the sequential (dotted line), simultaneous (dashed line), and combined escape (solid line) processes shown. $\mu = 10^{-5}$, $s_{del} = 10^{-2}$ and $s_{ben} = 10^{-1}$. N_{crit} is the value of N for which $T_{sim} = T_{seq}$. Monte Carlo simulation results T_{MC} (see text) in the absence of recombination are shown as filled circles for 1000 replicate realizations. Note that both axes are log-transformed.

$$N < \sim \frac{1}{2s_{del}}$$

T_{seq} is roughly

$$\frac{3}{2\mu}$$

because in this nearly neutral (Kimura 1968) range, fixation of valley chromosomes occurs almost as if by drift. The modest decline in T_{seq} seen in this range is the consequence of a growing $N\mu$ and the fact that the second term in equation (1) stops behaving neutrally before the first term does. It apparently differs from the similar decline previously observed (Kimura 1985a; Phillips 1996), which was essentially a deterministic response to very strong one-way mutational pressure. In the present case $\mu \ll s_{del}/10$, the boundary condition for this deterministic regime derived by Phillips (1996), and moreover simulation results (see below) demonstrate that the effect observed here is robust to back mutation. Thus, s_{del} determines the value of N at which T_{seq} begins its rapid increase. Similarly, for

$$N < \sim \frac{1}{2s_{ben}}$$

T_{sim} is roughly

$$\frac{s_{del}}{2\mu^2}$$

because here fixation of the AB chromosome occurs almost as if by drift, and so s_{ben} determines the value of N at which T_{sim} begins its rapid decline. Both functions are shifted down if μ increases (and up if μ decreases), but importantly must cross since we assume $s_{ben} \gg \mu$. Finally, for all parameter values, T_{esc} is proportional to $1/N$ for

$$N > \max\left(N_{crit}, \frac{1}{2s_{ben}}\right)$$

because in this region $u(N, s_{ben})$ is independent of N (Haldane 1927).

Values of T_{MC} (see Monte Carlo computer simulations) with no recombination are shown as filled circles in Figure 2. Only slight deviations from analytic expectations exist and these are of two sorts. In the sequential fixation regime back mutations (disregarded in eq. 1) inflate simulation results compared to expectation since after fixation of one of the mutationally intermediate valley chromosomes, populations may revert to the ab chromosome one or more times before finally fixing the AB chromosome. Additionally, in the simultaneous fixation regime simulations diverge from expectation as $2N$ approaches $1/\mu$, at which point the transit time through the population of the AB chromosome destined to fix becomes appreciable, violating the strong selection/weak mutation assumption. Decreasing μ or increasing s_{ben} reduces the fraction of time spent in this phase of the process and improves the fit (not shown).

To further confirm our understanding of the evolutionary dynamics in the two regimes, we recorded the minimum population mean fitness during the course of the simulations. For values of N in the sequential regime, this minimum was $1 - s_{del}$, confirming that one of the valley chromosomes was fixed in the population during the course of these simulations. For values of N in the simultaneous regime, mean fitness only dropped to $\sim 1 - 2\mu$, the Haldane-Muller mutational load expected for populations at mutation-selection equilibrium under genic selection (Crow and Kimura 1970; see also Discussion).

A good approximation for N_{crit} , the population size for which $T_{seq} = T_{sim}$ may be found by substituting

$$\frac{2s}{e^{4Ns} - 1}$$

for $u(N, -s)$ in the first term of equation (1) (justified for small s ; Kimura 1962), disregarding the second term in equation (1), which makes a vanishingly small contribution as N grows (Stephan 1996), and substituting $2s$ for $u(N, s)$ in equation (2) (justified for large N ; Haldane 1927). Together these simplifications yield

$$N_{crit} \approx \frac{\ln\left(\frac{s_{del}^2}{\mu \cdot s_{ben}} + 1\right)}{4s_{del}} \tag{4}$$

Values for this expression are shown in Figure 3 as a function of the compound variable $\mu \cdot s_{ben}$ for three values of s_{del} (10^{-1} , 10^{-3} , and 10^{-5} , solid, dashed, and dotted lines, respectively). This figure illustrates that the boundary between regimes is in the range of plausible sizes for natural populations, and that both regimes might thus be observed in nature.

Our treatment is for a fitness valley two mutations wide (Fig. 1). However the general form of the result, that when N exceeds some N_{crit} , T_{esc} becomes proportional to $1/N$, is robust to w , the number of mutations required to escape the local fitness peak. The rate of production of chromosomes differing by w mutations via the mutation of low fitness in-

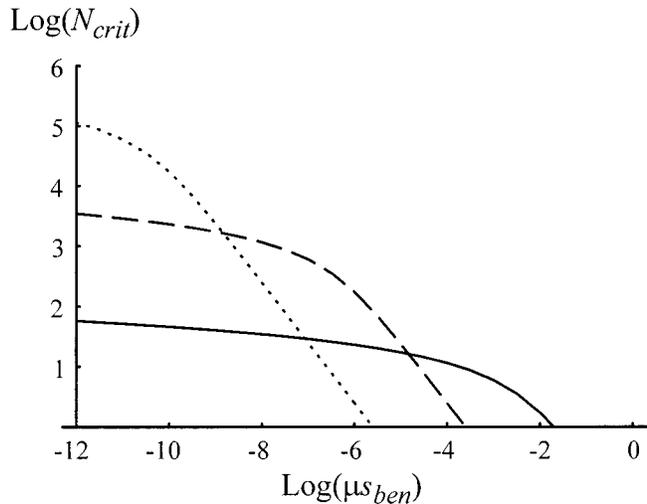


FIG. 3. N_{crit} , the population size that defines the boundary between the sequential and simultaneous regimes, as given by equation (4). Shown as a function of compound variable $\mu \cdot s_{ben}$ for $s_{del} = 10^{-1}$, 10^{-3} and 10^{-5} (solid, dashed, and dotted lines, respectively). Note that both axes are log-transformed.

intermediates maintained at mutation-selection equilibrium is proportional to μ^w (van Nimwegen and Crutchfield 2000), and so T_{sim} grows exponentially in w . Nevertheless, for fixed w , T_{sim} will maintain its linear dependence on $1/N$ for $N > 1/s_{ben}$, as will T_{esc} for N greater than the appropriate N_{crit} . However, to overcome the proportionality of T_{sim} to μ^{-w} , the intervening fitness valley may have to be substantially shallower or the net advantage across the valley may have to be substantially greater than the range of values in Figure 3 for N_{crit} to be at all small and the simultaneous fixation process across wider fitness valleys to be of great evolutionary importance in nature.

Genetic Recombination

Recombination is unlikely to influence the evolutionary escape from a local peak on the genotypic fitness landscape in populations operating in the sequential regime (Stephan 1996) as these rarely carry polymorphism at multiple loci simultaneously. However, as previously noted, recombination between loci may influence the simultaneous fixation process (Crow and Kimura 1965; Phillips 1996). In particular while AB chromosomes are still rare, most recombination events in which they are involved will be with the much more frequent ab chromosomes, producing low-fitness aB and Ab recombinants. Thus, recombination can extinguish nascent AB lineages that might otherwise reach fixation, slowing evolutionary escape from a local peak. In contrast, recombination can also create AB chromosomes if it occurs between aB and Ab chromosomes, thereby accelerating the escape process. Which effect dominates in any population is given approximately by the sign of linkage disequilibrium D : the former if $D = p_{ab}p_{AB} - p_{aB}p_{Ab} < 0$ and the latter otherwise, where p_{xy} is the frequency of chromosome xy . General analytic solutions for arbitrary recombination rate via the diffusion approximation have proven difficult (Kimura 1985a; Phillips

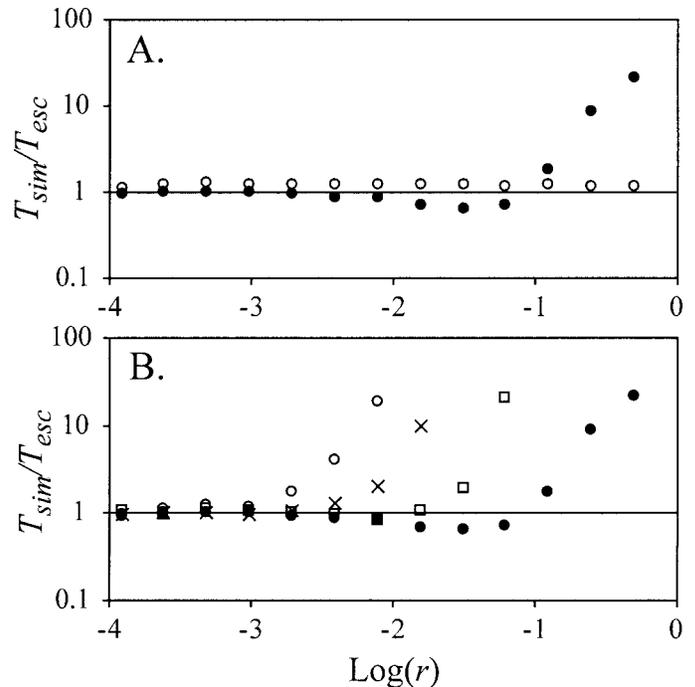


FIG. 4. Sensitivity to recombination rate r , for time to escape a local fitness peak in Monte Carlo simulations. (A) Results for populations in the sequential and simultaneous regimes: $N = 16$ (open circles) and $N = 1024$ (filled circles), respectively; other parameters as in Figure 2. (B) Results for populations in the simultaneous regime for $\log(s_{ben}) = -2.5$ (open circles), -2.0 (crosses), -1.5 (open squares), and -1.0 (filled circles). In all cases $N = 1024$ exceeds N_{crit} . For $\log(s_{ben}) < -1.0$ only 100 replicates were run. In both panels values shown are normalized by T_{esc} , the analytic expectation in the absence of recombination given by equation (3) and both axes are log-transformed.

1996; Stephan 1996), and here we follow a heuristic approach suggested by Monte Carlo simulation results.

Figure 4A represents T_{MC} normalized by T_{esc} (eq. 3) as a function of recombination rate r , for populations in the sequential (open circles) and simultaneous fixation (filled circles) regimes. Recombination has little influence on T_{MC} in the sequential regime as expected (Stephan 1996), because equation (1) assumes the absence of polymorphism, and T_{MC} is seen to exceed T_{esc} by a fixed proportion independent of r in this regime due to back mutations as described above. In contrast, recombination profoundly affects T_{MC} in the simultaneous fixation regime. In simulations lacking recombination, mean linkage disequilibrium before appearance of the AB chromosome that ultimately reached fixation is negative (not shown), suggesting that the marginal effect of recombination when recombination rates are small will be to initiate more AB lineages than are extinguished. This reasoning is consistent with the observation in Figure 4A that modest levels of recombination reduce T_{MC} in the simultaneous regime. As recombination rate increases further, mean linkage disequilibrium in simulations becomes positive (not shown), suggesting that recombination should now disrupt more AB chromosomes than it creates. Consistent with this picture, T_{MC} increases rapidly with recombination rate in this range. Much smaller effects of recombination on T_{MC} were

observed for smaller values of s_{del} (not shown), because weaker purifying selection has a much smaller influence on linkage disequilibrium (Stephan 1996).

Interestingly for $N > N_{crit}$, the time for evolutionary escape from a local peak in the presence of frequent recombination rises with N (not shown) because this process is again stochastic (Phillips 1996). However, for all parameter values explored in this regime, $T_{MC} < T_{seq}$, the expected time for the (stochastic) sequential process given by equation (1). This reflects the fact that these two stochastic processes differ: in the present case escape does not depend on the fixation of either low-fitness valley chromosome. Two observations (not shown) in our Monte Carlo simulations when $N > N_{crit}$ corroborate this point. First, for all recombination rates examined, minimum population mean fitness remained equal to $\sim 1-2\mu$, again reflecting the Haldane-Muller mutational load (see Discussion). Second, with frequent recombination the AB lineage that eventually reached fixation was not created by mutation of an Ab or aB chromosome, but rather by recombination between them. Thus, evolutionary escape in this regime depends only on the transient, simultaneous segregation of both low-fitness valley chromosomes at sufficient frequency that recombination can generate AB chromosomes, and this drift-mediated component of the escape process accounts for the positive relationship between T_{MC} and N in this regime.

Figure 4A demonstrates that a threshold recombination rate exists above which the deterministic simultaneous fixation process gives way to a stochastic simultaneous fixation process. Motivated by these simulation results we sought an expression for this critical recombination rate (Crow and Kimura 1965). Recall that a beneficial mutation of genic selective effect s is lost to genetic drift in populations of arbitrary size with probability $1-2s$ (Haldane 1927) because substantial Wright-Fisher sampling variance operates while the beneficial mutation exists in only a small integer number of copies in the population. However, once such a mutation increases in frequency and there are more than roughly $1/s$ copies in the population (Gale 1990), stochastic loss of one lineage is unlikely to extinguish the mutation's ultimate transit to fixation. This reasoning has led others to divide the fixation process for a beneficial mutation into a stochastic and a deterministic phase (Gale 1990; Gillespie 1991). We conceptualized the loss of an AB chromosome to recombination as analogous to its loss to drift, implying that recombination is only likely to extinguish the entire subpopulation of AB chromosomes if it occurs while the frequency of an AB chromosome otherwise destined to fixation remains in the stochastic phase.

Beneficial mutations of selective effect s destined for fixation remain in the stochastic phase for approximately $-[1.27 + \ln(s)]/s$ generations (Gale 1990). If the rate of recombination per individual is r then the mean number of generations between recombination events in any lineage is $1/r$ (Crow and Kimura 1965), suggesting that as long as the recombination rate between loci is less than

$$r_{crit} \approx \frac{-s_{ben}}{1.27 + \ln(s_{ben})}, \quad (5)$$

recombination is unlikely to disrupt nascent AB lineages oth-

erwise destined to fixation. To test this prediction additional Monte Carlo computer simulations were run and very close coincidence to this expectation was observed as shown in Figure 4b. For $\log(s_{ben})$ equal to -1.0 , -1.5 , -2.0 , and -2.5 , the log-transformed values of r_{crit} are -1.01 , -1.84 , -2.52 , and -3.15 respectively (the x -axis of Fig. 4 is similarly log-transformed); these are very nearly exactly the recombination rates at which T_{MC} first exceeds T_{esc} .

One might suppose that functional interactions between mutations such as those illustrated in Figure 1 will more commonly occur within individual genes (e.g., Stephan and Kirby 1993; Kondrashov et al. 2002) and recent empirical estimates give recombination rates in metazoans as roughly 1 cM/Mb (Hey and Kliman 2002; Jensen-Seaman et al. 2004). Thus, if $N > N_{crit}$ a pair of jointly beneficial but individually deleterious mutations 10^5 nucleotides apart (i.e., within a typical protein-coding gene with introns) will deterministically fix simultaneously if $s_{ben} > \sim 0.004$, a seemingly small selective advantage. The deterministic fixation of larger epistatically beneficial effects between correspondingly more distant mutations will similarly be likely to survive recombination.

DISCUSSION

The evolutionary dynamics that describe population escape from a local peak on the genotypic fitness landscape such that as illustrated in Figure 1 are sensitive to both population size, N (Gillespie 1984; Carter and Wagner 2002; Iwasa et al. 2004) and to recombination rate, r (Crow and Kimura 1965; Phillips 1996). We show that three evolutionary regimes exist: for small N escape is likely to be via the stochastic sequential fixation of individual mutations. For large N escape is likely to be via the simultaneous fixation of the jointly beneficial mutations, although for small r this process is deterministic while for large r it is again stochastic. Using an intuitively simple analytic model we provide expressions for the critical values for these parameters that define the boundaries between regimes, which together suggest that all three regimes are likely to operate in nature. Because the evolutionary escape dynamics in the first and third regimes are stochastic, expected times are inversely correlated with population size, whereas the expected time for populations in the second, deterministic regime is positively correlated with size. Thus, perhaps surprisingly, evolutionary escape from some local peaks on the genotypic fitness landscape in natural populations may be fastest in large populations.

Prior Art

The contrast between the small N and large N /small r regimes has previously been noted. For $s_{ben} = 0$ what we term the simultaneous fixation process still operates (Kimura 1985b; Stephan 1996; Innan and Stephan 2001), although here transit across the fitness valley is neutral. Because

$$u(N,0) = \frac{1}{2N},$$

our equation (2), becomes insensitive to N as is expected for the time to fix a genotype with no net selective effect (Stephan 1996). Thus, whereas the dashed line in Figure 2 (T_{sim} for

$s_{ben} > 0$) declines as N increases, the analogous function for $s_{ben} = 0$ is a horizontal line with the same y -intercept, making clear graphically that setting $s_{ben} = 0$ increases both N_{crit} and T_{sim} .

As noted the historical antecedents for the case $s_{ben} > 0$ are even older, beginning with Gillespie (1984). Although he recognized that the expected time for the simultaneous fixation process drops as population size increases, he felt that the proportionality to $1/\mu^2$ might make this unrealistically slow. Perhaps in part to address this criticism, Carter and Wagner (2002) proposed a diploid model in which the deleterious mutations are recessive and thus mutationally intermediate alleles are hidden from selection while heterozygous, causing them both to fix more rapidly and to segregate at higher frequency. This accelerates both processes (a similar effect of dominance was also observed by Phillips 1996), although we show explicitly that dominance is not necessary for either. No analog of N_{crit} in the presence of dominance was presented by Carter and Wagner (2002), but examination of their figure 5 suggests that for a wide range of parameter values the effects of dominance balance in the two processes, and that our expression appears to give a reasonable approximation for the boundary between regimes in its presence. More recently it has been suggested that the simultaneous process may be very important in the context of the somatic evolution of cancer (Nowak et al. 2002; Iwasa et al. 2004), especially as genetic instabilities (Lengauer et al. 1998) can inflate the mutation rate from Ab and aB genotypes by many orders of magnitude (Iwasa et al. 2004). This accelerates only the simultaneous fixation process by this same factor and thereby reduces N_{crit} .

We note that while the expected time to cross any given fitness valley in the large N /small r regime may be long, one might imagine that evolving populations simultaneously sample low-fitness variants segregating in a great many selectively independent fitness valleys, and that they would thus cross some such valley with modest frequency. Moreover, by equating equation (2) with the y -intercept of equation (1) one finds that for N larger than $\sim s_{del}/12\mu \cdot s_{ben}$ the simultaneous process becomes faster than the sequential process for any N (in spite of the latter's proportionality only to $1/\mu$), suggesting that sufficiently shallow fitness valleys will be crossed most quickly via the former process in all population sizes.

The contrast between large N /small r and large N /large r dynamics has also been long recognized. Crow and Kimura (1965) explored this question in an infinite population model and give $s_{ben}/(s_{ben} + 1)$ as the critical recombination rate defining the boundary between these deterministic and stochastic regimes, which gives log-transformed critical values of -1.04 , -1.51 , -2.00 , and -2.50 for the four values of s_{ben} employed in Figure 4B. This expression assumes that a subpopulation of AB chromosomes persists stably, and represents the critical recombination rate below which its size increases selectively in spite of recombination. Examination of Figure 4B shows that this approach overestimates the upper recombinational boundary of the deterministic regime when $s_{ben} < \sim 0.1$, and that in this case a more accurate picture of the process is one in which AB lineages transiently

appear but may be extinguished (Gale 1990; Gillespie 1991) while still rare if recombination is sufficiently common.

Sewall Wright was the first to explicitly make the point that epistasis for fitness could result in mutationally isolated peaks on the genotypic fitness landscape (Wright 1930, 1932; Provine 1986). He posed the problem of evolutionary fitness increase for a population fixed for a genotype corresponding to a local peak (Wright 1932), engendering a considerable literature involving models of population subdivision predicated on the stochastic sequential escape process (Wright 1932; Wade and Goodnight 1991; Moore and Tonsor 1994; Coyne et al. 2000; Goodnight and Wade 2000). In this connection, we note that the consequences of population structure under both the deterministic and stochastic simultaneous escape processes remain largely unexplored.

Modeling Evolution on Fitness Landscapes with Epistasis

Interest in fitness landscapes stems from the desire to develop intuition into the evolutionary behavior of populations in the presence of arbitrary epistasis (Wright 1932). Although the genotypic fitness landscape (e.g., Fig. 1) correctly represents all epistatic interactions acting in a genome, we have seen that naive evolutionary inference developed in this framework can be misleading. In particular, the implicit assumption that a population can be represented by a single point in genotypic sequence space disregards the possibility that polymorphism can play a critical role in its evolutionary dynamics. Consequently, sufficiently large populations may appear to become decoupled from the contours of the genotypic fitness landscape (Stephan 1996), and we have seen that they can bridge fitness valleys with no substantial reduction in mean fitness (Iwasa et al. 2004). These considerations illustrate that in the presence of epistasis, accurately modeling evolution on the genotypic fitness landscape may require explicitly representing the population as a distribution of points in genotypic sequence space. Such an approach accommodates segregating polymorphism and moreover makes clear why the population mean fitness need not track the underlying genotypic fitness landscape precisely.

Perhaps a more commonly employed strategy for modeling the evolution of polymorphic populations has been to track allele frequencies at all loci. Mapping each such population composition to its mean fitness defines the population mean fitness landscape (e.g., Wright 1982; reviewed in Provine 1986; Gavrilets 2004), and in each generation an evolving population will adjust allele frequencies in such a way as to maximize the incremental increase in mean fitness (Fisher 1930). Note, however, that in the presence of epistasis (precisely the biological possibility that motivates interest in fitness landscapes), population mean fitness is also a function of linkage disequilibrium D . Because this framework offers no insight into the dynamics of D , it appears to be of limited predictive utility. The present results illustrate this point. The population mean fitness landscape corresponding to Figure 1 exhibits a ridge spanning the fitness valley if and only if D is sufficiently positive. Our observation that mean fitness declines only very slightly during evolutionary escape when $N > N_{crit}$ therefore suggests that evolutionary escape from the local peak in large populations must coincide with a tran-

sition in *D*. However the model does not clarify the determinants of this event. (In fact, selection on *AB* chromosomes will quickly drive *D* strongly positive after their appearance; recombination reduces $|D|$, repeatedly eroding the ridge and slowing escape.) Thus, although a post hoc description of the escape process is possible in the context of the population mean fitness landscape, in the presence of arbitrary epistasis this model appears to offer little predictive power.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of A. Berry, C. Burch, D. Hartl, R. Kulathinal, R. C. Lewontin, C. Meiklejohn, P. C. Phillips, A. Poon, O. Silander, O. Tenaillon, Y. Tao, G. Wagner, R. Watson, and two anonymous reviewers. This work was supported in part by grant DEB-0343598 from the National Science Foundation to Daniel L. Hartl and DMW and a grant from the National Institutes of Health (NIGMS) to LC.

LITERATURE CITED

- Burch, C. L., and L. Chao. 2000. Evolvability of an RNA virus is determined by its mutational neighborhood. *Nature* 406: 625–628.
- Carter, A. J. R., and G. P. 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.
- Coyne, J. A., N. H. Barton, and M. Turelli. 2000. Is Wright's shifting balance process important in evolution? *Evolution* 54: 306–317.
- Crow, J. F. 1991. Was Wright right? *Science* 253:973.
- Crow, J. F., and M. Kimura. 1965. Evolution in sexual and asexual populations. *Am. Nat.* 99:439–450.
- . 1970. An introduction to population genetics theory. Harper and Row, New York.
- Ewens, W. J. 1979. *Mathematical population genetics*. Springer, Berlin.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford, U.K.
- Gale, J. S. 1990. *Theoretical population genetics*. Unwin Hyman, London.
- Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton Univ. Press, Princeton, NJ.
- Giachetti, C., and J. J. Holland. 1988. Altered replicase specificity is responsible for resistance to defective interfering particle interference of an Sdi⁻ mutant of vesicular stomatitis virus. *J. Virol.* 62:3614–3621.
- Gillespie, J. H. 1984. Molecular evolution over the mutational landscape. *Evolution* 38:1116–1129.
- . 1991. *The causes of molecular evolution*. Oxford Univ. Press, Oxford, U.K.
- . 1993. Substitutional processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. *Genetics* 134:971–981.
- Goodnight, C. J., and M. J. Wade. 2000. The ongoing synthesis: a reply to Coyne, Barton and Turelli. *Evolution* 54:317–324.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. V. *Proc. Camb. Philol. Soc.* 23:838–844.
- Hey, J., and R. M. Kliman. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* 160:595–608.
- Innan, H., and W. Stephan. 2001. Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. *Genetics* 159:389–399.
- Iwasa, Y., F. Michor, and M. A. Nowak. 2004. Stochastic tunnels in evolutionary dynamics. *Genetics* 166:1571–1579.
- Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C.-F. Chen, M. A. Thomas, D. Haussler, and H. J. Jacob. 2004. Comparative recombination rates in the rat, mouse and human genomes. *Genome Res.* 14:528–538.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- . 1968. Evolutionary rate at the molecular level. *Nature* 217: 625–626.
- . 1985a. Diffusion models in population genetics with special reference to fixation time of molecular mutants under mutational pressure. Pp. 19–39 in T. Ohta and K. Aoki, eds. *Population genetics and molecular evolution*. Springer, Berlin.
- . 1985b. The role of compensatory neutral mutations in molecular evolution. *J. Genet.* 64:7–19.
- Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. USA* 99:14878–14883.
- Lengauer, C., K. W. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. *Nature* 396:643–649.
- Maisnier-Patin, S., O. G. Berg, L. Lijas, and D. I. Andersson. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol. Microbiol.* 46: 355–366.
- Maynard Smith, J. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–565.
- Moore, F. B.-G., and S. J. Tonsor. 1994. A simulation of Wright's shifting-balance process: migration and the three phases. *Evolution* 48:69–80.
- Nowak, M. A., N. L. Komarova, A. Sengupta, P. V. Jallepalli, I.-M. Shih, B. Vogelstein, and C. Lengauer. 2002. The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci. USA* 99:16226–16231.
- O'Hara, P. J., F. M. Horodyski, S. T. Nichol, and J. J. Holland. 1984. Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5'-terminal and fewer 3'-terminal mutations in a stepwise manner. *J. Virol.* 49:793–798.
- Phillips, P. C. 1996. Waiting for a compensatory mutation: phase zero of the shifting-balance process. *Genet. Res.* 67:217–283.
- Provine, W. B. 1986. *Sewall Wright and evolutionary biology*. Univ. of Chicago Press, Chicago.
- Schrag, S. J., V. Perrot, and B. R. Levin. 1997. Adaptation to the fitness cost of antibiotic resistance in *E. coli*. *Proc. R. Soc. Lond. B* 264:1287–1291.
- Stephan, W. 1996. The rate of compensatory evolution. *Genetics* 144:419–426.
- Stephan, W., and D. A. Kirby. 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. *Genetics* 135:97–103.
- van Nimwegen, E., and J. P. Crutchfield. 2000. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bull. Math. Biol.* 62:799–848.
- Wade, M. J., and C. J. Goodnight. 1991. Wright's shifting balance theory: an experimental study. *Science* 253:1015–1018.
- Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* 26:601–629.
- Wright, S. 1930. Review of the genetical theory of natural selection. *J. Hered.* 21:349–356.
- . 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Pp. 356–366 in D. F. Jones, ed. *Proceedings of the sixth international congress of genetics*. Brooklyn Botanic Garden, Menasha, WI.
- . 1965. Factor interaction and linkage in evolution. *Proc. R. Soc. Lond. B* 162:80–104.
- . 1970. Random drift and the shifting balance theory of evolution. Pp. 1–31 in K.-I. Kojima, ed. *Mathematical topics in population genetics*. Springer, New York.
- . 1982. Character change, speciation and the higher taxa. *Evolution* 36:427–443.

Corresponding Editor: P. Phillips