

# Scaling expectations for the time to establishment of complex adaptations

Michael Lynch<sup>1</sup>

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Michael Lynch, August 11, 2010 (sent for review March 16, 2010)

Although the vast majority of research in evolutionary biology is focused on adaption, a general theory for the population-genetic mechanisms by which complex adaptations are acquired remains to be developed. The issue explored here is the procurement of novel traits that specifically require multiple mutations to achieve a fitness advantage. By highlighting the roles played by the forces of mutation, recombination, and random genetic drift, and drawing from observations on the joint constraints on these factors, the ways in which rates of acquisition of specific types of adaptations scale with population size are explored. These general results provide insight into a number of ongoing controversies regarding the molecular basis of adaptation, including the adaptive utility of recombination and the role of drift in the passage through adaptive valleys.

adaptive evolution | complex traits | evolutionary rate | molecular evolution | recombination

Recent empirical observations imply that approximate scaling laws exist for several fundamental evolutionary forces (1–3). First, across the full domain of life, there is an inverse relationship between population density and organism size. This relationship bears importantly on the power of random genetic drift, which is expected to scale negatively (although not necessarily linearly) with total population size (4). Second, the recombination rate per physical distance on chromosomes scales inversely with genome size. This feature is a simple consequence of an apparent structural constraint across all sexually reproducing eukaryotes—the occurrence of approximately one crossover event per chromosome arm per meiosis, and the fact that most variation in genome size is associated with variation in chromosome size rather than chromosome number. Finally, the mutation rate per nucleotide site per generation scales negatively with the genetic effective size of a population, possibly because lineages more vulnerable to random genetic drift are less efficient at maintaining high-fidelity replication/repair machinery (3, 5). The fact that all three of these nonadaptive forces of evolution influence the efficiency of selection raises the question as to whether general scaling laws also exist for the exploitation of various pathways to adaptive evolution.

The development of theory in this area is rendered difficult by the multidimensional nature of the problem. One strategy has been to ignore all deleterious mutations and to assume that selection is strong enough and mutation weak enough relative to the power of random genetic drift and recombination that evolution always proceeds by the sequential fixation of single mutations (e.g., refs. 6–11). Such an approach provides a useful entree into the evolutionary dynamics of rare adaptive mutations with large effects. Under these conditions, the expectations are clear—with larger numbers of mutational targets and a reduced power of random genetic drift, the rate of adaptation will increase with population size, although more slowly than expected under the assumption of sequential fixation (12, 13). The motivation for these models, which are specifically focused on total organismal fitness, derives from case studies of adaptations with apparently simple genetic bases, e.g., some aspects of insecticide resistance (14), skin pigmentation (15), and skeletal morphology in vertebrates (16).

Nevertheless, a broad subset of adaptations cannot be accommodated by the sequential model, most notably those in which

multiple mutations must be acquired to confer a benefit. Such traits, here referred to as complex adaptations, include the origin of new protein functions involving multiresidue interactions, the emergence of multimeric enzymes, the assembly of molecular machines, the colonization and refinement of introns, and the establishment of interactions between transcription factors and their binding sites, etc. The routes by which such evolutionary novelties can be procured include sojourns through one or more deleterious intermediate states. Because such intermediate haplotypes are expected to be kept at low frequencies by selection, evolutionary progress would be impeded in large populations where sequential fixation is the only path to adaptation. However, in all but very small populations, complex adaptations appear to be achieved by the fortuitous appearance of combinations of mutations within single individuals before fixation of any intermediate steps at the population level (e.g., refs. 17–26).

The goals of the following work are twofold. First, although most theory on the evolution of complex adaptations has been focused on nonrecombining systems, because recombination can serve as a creative force in evolution (27, 28), prior results might greatly underestimate the rate of emergence of novel adaptations. It is shown here, however, that recombination often plays a fairly minor role in the rate of acquisition of complex adaptations, except in a narrow range of chromosomal positions, and in some cases is inhibitory. Second, to understand the implications of the general theoretical results, there is a need to confine the analyses to the known parameter space of the key underlying factors. Here, the use of the empirically determined relationship between the mutation rate and the power of drift is used to explore how the lability of alternative molecular paths to adaptation scales with aspects of the population-genetic environment.

## Model and Results

### Recombination and the Rate of Emergence of Complex Adaptations.

Consider the situation in which two specific mutations must be acquired to achieve the novel adaptive state. Starting from state **ab**, the first step involves the production of **aB** or **Ab** intermediates. Initially, recombination will have no impact on the evolutionary dynamics, as heterozygotes carrying intermediate-state and ancestral **ab** haplotypes will yield no new daughter products. However, if both types of intermediate haplotypes simultaneously rise to moderate frequencies, recombination between them can produce the adaptive **AB** combination. The effective size of a population ( $N_e$ ) determines the likelihood of such cooccurrence.

If  $N_e$  is sufficiently small that the waiting times for new mutations are long compared with the fixation times, the **Ab** and **aB** haplotypes will essentially never encounter each other, and the only route to the final adaptation will be the sequential fixation of the **A** and **B** mutations. Letting  $u$  be the rate of mutation at each site, this sequential mode of evolution predominates only if  $N_e \ll 1/\sqrt{8u}$ , a fairly restrictive domain given that  $u > 10^{-10}$  for

Author contributions: M.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

<sup>1</sup>E-mail: milynch@indiana.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010836107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010836107/-DCSupplemental).

all of cellular life (1, 3). At larger population sizes, recombination within **ab/Ab** individuals will often create **AB** gametes at rates beyond those expected by secondary mutation, as the recombination rate per nucleotide site is typically on the order of the mutation rate (2). However, in the initial stages of this process, because the **ab** gamete still has high frequency, newly arisen **AB** haplotypes will also be subject to recombinational loss. If the intermediate haploid states are neutral, mutation pressure can eventually drive the **aB** and **Ab** types to high enough frequencies that recombinational production of the **AB** haplotype outpaces the rate of loss. However, if the intermediate states are deleterious, selection can permanently maintain the **Ab** and **aB** haplotypes at low enough frequencies to prevent progression through a sufficiently strong recombination barrier. This inhibitory condition occurs when the rate of promotion of the **AB** type by selection ( $s$  denoting the selective advantage) is exceeded by the rate of loss by recombination.

These verbal arguments are now developed in a more formal fashion by evaluating the mean time to establishment of the **AB** haplotype (mutational origin plus fixation,  $\bar{t}_c$ ). Diploidy is assumed throughout, although the general expressions still pertain to haploids with a sexual phase, provided  $N_c$  is substituted for  $2N_c$ . The rate of recombination between sites per generation is denoted as  $c$ .

**Neutral Intermediates.** In the low recombination domain ( $0 \leq c < s$ ), the mean number of generations until establishment of the double mutant can be approximated as the reciprocal of the sum of arrival rates of adaptive alleles destined to fixation by three different pathways,

$$\bar{t}_c \approx \frac{1}{r_s + r_t + r_r}, \quad [1]$$

where  $r_s$  is the rate of establishment by sequential fixation (i.e., fixation of the **Ab** or **aB** state followed by mutation to **AB**),  $r_t$  is the rate of stochastic tunneling without recombination (i.e., creation of a successful **AB** haplotype by secondary mutation before fixation of an intermediate state), and  $r_r$  is the rate of establishment initiated by recombination events. This approximation follows from the behavior of independent, exponentially distributed random variables, although the three paths are not entirely independent. The additional time to fixation of the final adaptive allele is ignored, as this time is generally negligible relative to the arrival times of fixation-destined mutations. Expanding from prior work, under conditions of intermediate-state neutrality,  $r_s \approx 2u/[1 + (1/2Ns)]$  and  $r_t \approx (1 - e^{-4Nu})\sqrt{u}\phi_f$ , where  $\phi_f = [1 - e^{-2(s-c)(N_c/N)}]/[1 - e^{-4N_c(s-c)}]$  is the probability of fixation of a newly arisen **AB** haplotype, with  $N$  being the absolute population size (26).

The rate of establishment of the **AB** haplotype by recombinational production ( $r_r$ ) can be approximated by using a branching-process approach. Imagine a newly arisen neutral first-step mutation of type **Ab**. Conditional on survival to generation  $t$ , the probability of which is  $\sim 2/t$  (ref. 29, Chap. 5), such a haplotype will have reached an average number of  $t/2$  copies (because the expected number is always 1.0). Having arisen to such a level, it will then take an average  $t$  generations to drift to zero in the absence of any advantageous secondary mutations. Thus, with an average  $t/4$  copies over this time span of  $2t$  generations, with the alternative **aB** first-step mutation having an expected frequency of  $\sim ut$  at time  $t$  (as a result of cumulative mutation pressure), and with recombination between **aB** and **Ab** haplotypes creating **AB** alleles at rate  $c/2$ , the rate of establishment of the adaptive allele by the recombinational path is expected to be proportional to  $(1 - e^{-4Nu})(cu\phi_f/4)^{1/3}$ , where the first term in parentheses denotes the probability of arrival of at least one intermediate-state allele per generation.

This scaling of  $r_r$  with the cube root of  $c$  is consistent with earlier work on the arrival time of the first recombinant (30). However, these authors did not consider the additional matter of

fixation and after applying several approximations concluded that the arrival rate for first recombinants scales with  $(cN_cu^2)^{1/3}$  in contrast to the overall behavior suggested above. It is shown below that these two results are reconciled quite closely with the single formula,

$$r_r \approx (4/3)(1 - e^{-4Nu}) \left( \frac{4N_c c u^2 \phi_f}{1 + 32N_c u} \right)^{1/3}, \quad [2]$$

with a shift in behavior from that suggested by Christiansen et al. (30) to that suggested above as  $4N_c u$  progresses beyond 1/8.

The overall influence of recombination on  $\bar{t}_c$  is a function of the two opposing ways in which the **AB** haplotype is influenced—the rate of origin of **AB** gametes by recombination within doubly heterozygous (**aB/Ab**) parents is proportional to  $c$ , whereas the net selective advantage of the resultant **AB** haplotypes is discounted to  $s - c$  by subsequent recombinational breakdown (whereas **ab** haplotypes still predominate). Thus, because the product  $c(s - c)$  is maximized at  $c = s/2$ , two-site adaptations with neutral intermediate steps are expected to emerge most rapidly in chromosomal settings where the recombination rate is half the selective advantage of the final adaptation.

The maximum benefit of recombination can be determined by comparing  $(r_t + r_r)$  when  $c = s/2$  with the rate of tunneling in the absence of recombination at the large population-size limit, the ratio of which is  $\approx \{1 + (2/3)[sN/(2N_c u)]^{1/6}\}/\sqrt{2}$ . Assuming  $N_c = N$ , with a microbial mutation rate on the order of  $u = 10^{-10}$ , this ratio becomes 4.9, 6.9, and 9.8 for  $s = 0.0001$ , 0.001, and 0.01, whereas with a mutation rate more typical of a multicellular eukaryote,  $\sim 10^{-8}$ , the ratios are 2.7, 3.6, and 4.9. Actual population sizes are generally much greater than effective population sizes (2, 4), but the preceding values would be inflated only by a factor of 3 if  $N/N_c$  were as large as 1,000. Thus, as these calculations give the extreme when  $c = s/2$ , the influence of recombination on the rate of establishment of a two-site adaptation with neutral intermediates will usually be much less than an order-of-magnitude effect.

For the high-recombination domain ( $c > s$ ), the gamete frequencies will be maintained near linkage equilibrium (except for the influence of drift), in which case selection will operate on each component in an effectively additive manner (31). In this case, for sufficiently small population sizes ( $4Nu < 1$ ),  $\bar{t}_c$  is adequately expressed by Eq. 1 with  $r_r = 0$  and  $\phi_f = [1 - e^{-2s(N_c/N)}]/[1 - e^{-4N_c s}]$ . For the limiting case in which the population is large enough to be treated in a deterministic fashion (with the population-wide rate of introduction of first-step alleles  $4Nu > 1$ ), the mean time to establishment can be shown to be

$$\bar{t}_c \approx \frac{\arctan(\sqrt{s/u})}{\sqrt{us}} \quad [3a]$$

generations. This time reduces to  $\bar{t}_c \approx 1.57/\sqrt{us}$  provided  $\sqrt{s/u} > 10$ , which will generally be the case for the types of adaptation under consideration. Under the same conditions, but in the absence of recombination,

$$\bar{t}_c \approx \frac{1}{2u\sqrt{N_c s}} \quad [3b]$$

(26). Thus, as the ratio of the times given by Eqs. 3a and Eqs. 3b is  $1.57\sqrt{4N_c u}$ , recombination slows the rate of establishment of a two-site adaptation via neutral intermediates by a factor  $< 1.57 \times$  the square root of the effective number of intermediate-state alleles arising per generation. The transition to the high-recombination behavior defined by Eq. 3b is almost immediate and nearly independent of the recombination rate once  $c$  exceeds  $s$  (Fig. S1).

**Deleterious Intermediates.** Provided the power of drift ( $1/2N_e$ ) is sufficiently weaker than the selective disadvantage of intermediate **Ab** and **aB** haplotypes ( $\delta$ ), prior to positive selection for a double mutant, deleterious first-step alleles will be maintained by selection–mutation balance at expected frequencies  $\sim u/\delta$ . These maladapted haplotypes will then serve as launching pads for second-step adaptive mutations. Iwasa et al. (32) derived an iterative branching-process expression for the probability of first arrival of an **AB** combination in the absence of recombination, which, after incorporating the fixation probability, yields a closed-form approximation for the rate of stochastic tunneling,

$$r_t \approx \frac{8(N_e u)^2 s}{N \delta} \quad [4a]$$

(26). With the rate of establishment by sequential fixation (typically of negligible importance when  $4N_e \delta > 1$ ) being given by equation 4a in ref. 26, the mean time to establishment when  $r_r = 0$  can again be estimated by applying Eq. 4a to Eq. 1.

The role of recombination can be included by noting that for populations with sufficiently large  $N_e$  to maintain the two intermediate-state alleles at stable selection–mutation balance, the expected frequency of double heterozygotes is  $2(u/\delta)^2$ . With each such individual producing an average of two successful gametes, a fraction  $c/2$  of which are **AB**, the effective rate of recombinational production of the novel **AB** allele is then  $2N_e c (u/\delta)^2$  per generation. Assuming  $s > c$ , with the fixation probability of **AB** alleles being  $\approx 2(s - c)(N_e/N)$  at large  $N_e$ , the total rate of establishment becomes

$$r_t + r_r \approx \frac{4(N_e u)^2 (s - c)}{N \delta} \left(2 + \frac{c}{\delta}\right). \quad [4b]$$

The ratio of rates of establishment with and without recombination in large populations is  $\sim (s - c)[2 + (c/\delta)]/(2s)$ , and provided  $s > c$ , we can again anticipate that the rate of advancement of the adaptive combination will be maximized when  $c \approx s/2$ , in which case the ratio of rates reduces to  $0.5[1 + (s/4\delta)]$ . Thus, unless  $s \gg \delta$ , the effect of recombination is expected to be small. These results are inconsistent with the conclusion obtained under the assumption of an infinite population size, where recombination always prevents fixation of the double mutant (33).

To obtain the behavior of  $\bar{t}_c$  over the full range of population sizes, it must be recognized that at sufficiently small  $N_e$  recombination will be so uncommon as to make a negligible contribution to adaptational advance or inhibition. However, because at intermediate population sizes some stochastic paths to establishment will involve recombination and others will not, an approximation for the overall rate of establishment is

$$r_T = e^{-kNu} r_{T,0} + (1 - e^{-kNu}) r_{T,r}, \quad [5]$$

where  $r_{T,0}$  and  $r_{T,r}$  are the total expected rates in the absence and the presence of recombination (given by Eq. 4a and Eq. 4b). The exponential term  $kNu$ , which is proportional to the expected number of gametes carrying intermediate-state alleles at selection–mutation equilibrium, must exceed 0.1 for recombination to play a significant role. However, the exact definition of  $k$  has not been forthcoming, and in the following section, its numerical value is simply obtained by inspection of the rate of transition between the nonrecombination ( $c = 0$ ) and recombination ( $0 < c < s$ ) regimes of behavior in computer simulations.

Finally, in contrast to the situation in which  $s > c$ , if the rate of recombination exceeds the selective advantage of the **AB** haplotype, recombinational breakdown to deleterious intermediates will present an extremely strong barrier to establishment of the **AB** type. This result is because almost all recombinational events involving a newly arisen **AB** haplotype will have an **ab** partici-

pant, generating the maladaptive **Ab** and **aB** products. Only if the power of drift substantially exceeds the selective disadvantage of the intermediate haplotypes,  $4N_e \delta < 1$ , will the intermediate-state haplotypes ever drift to high enough frequencies,  $\approx \delta/(s + 2\delta)$ , for a newly emergent **AB** allele to overcome the recombinational barrier. The probability that a newly arisen underdominant haplotype proceeds to fixation is

$$\phi_r \approx \frac{(2/N)e^{-N_e \delta/\alpha} \sqrt{N_e \delta \alpha}}{\sqrt{\pi} \operatorname{erf}\{\sqrt{N_e 2\delta \alpha} [1 - (0.5/\alpha)]\} + \operatorname{erf}[\sqrt{\delta N_e/\alpha}]}, \quad [6]$$

where  $\operatorname{erf}(x)$  is the error function evaluated at  $x$ , and  $\alpha = (s + 2\delta)/(2\delta)$  (34). Noting that both the **A** and **B** alleles must elevate to high frequency for the **AB** adaptation to take hold, the rate of establishment under effectively free recombination in large populations becomes

$$r_r \approx (4Nu\phi_r)^2/\delta, \quad [7]$$

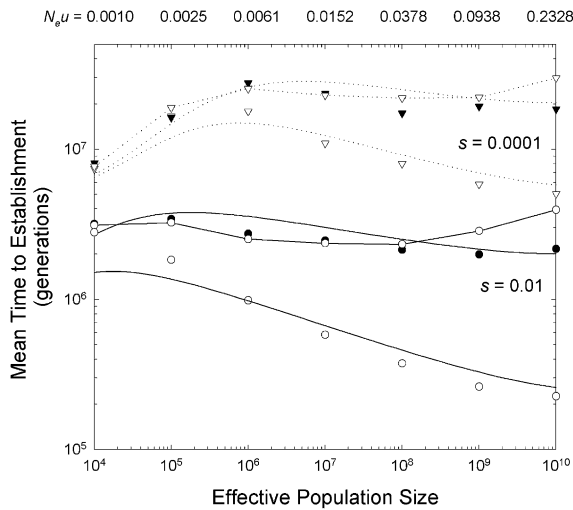
which accounts for the fact that one-step mutations arise at rate  $4Nu$  per generation and are retained for an average of  $1/\delta$  generations of exposure to potential alleles at the alternative locus. The mean time to establishment is then  $\approx 1/(r_s + r_r)$ .

**Scaling of the Time to Establishment with the Effective Population Size.** Although the preceding results provide general guidance as to how mutation, recombination, and random genetic drift influence  $\bar{t}_c$ , any attempt to derive a biologically meaningful theory for the ability of populations to exploit various routes to adaptation ought to account for the fact that these three parameters are not independently distributed. Most notably, per-generation mutation rates increase with the power of random genetic drift. The base-substitution mutation rate per nucleotide site (in units of  $10^{-9}$  per generation) scales as  $6.12N_e^{-0.60}$  (with  $N_e$  in units of  $10^6$  individuals) (3). Using this relationship, the mutation rate will now be treated as a dependent variable in evaluating how the times to establishment of novel adaptations are likely to scale with  $N_e$  in natural populations. To determine the precision of the preceding analytical approximations, comparisons are made with simulation results for a Wright–Fisher diploid population (with  $N_e = N$ ), using previously developed computational machinery (26).

For the situation in which the intermediate steps to adaptation are neutral, the theory developed above generally provides a close fit to the simulated data (Fig. S1). Due to the partially compensating effects of the increased mutation rate with the decline in population size, the mean time to establishment of a novel two-step allele is expected to be nearly independent of the effective population size (scaling as  $N_e^{-0.1}$ , provided  $4N_e s > 1$ ) (Fig. 1). This result follows by noting for the neutral-intermediate two-site model that the rate of tunneling in the absence of recombination is proportional to  $\sim N_e u \sqrt{u}$  and substituting  $u \propto N_e^{-0.6}$ . Such weak scaling is a strong departure from the view that would be obtained if the mutation rate were treated as a constant, in which case  $\bar{t}_c$  scales as  $N_e^{-1.0}$  provided  $4N_e u < 1$ .

As anticipated from the results described above, recombination rates equal to  $\sim s/2$  can result in a severalfold increase in the rate of adaptation when  $s$  is large and  $\delta = 0$ , although the effect asymptotically approaches zero in small populations due to the rarity of simultaneously segregating polymorphisms (Fig. 1). Provided  $c < s$ , recombination does not greatly alter the scaling of  $\bar{t}_c$  with  $N_e$ , as Eq. 2 shows that in this case the scaling of the recombinational contribution to  $\bar{t}_c$  is  $\propto N_e^{-0.2}$ . In very large populations, the rate of establishment is moderately impeded when  $c > s$ .

The models developed above also fit the simulated data with deleterious intermediates quite well (Fig. S2). In this case, in contrast to the situation with neutral intermediates,  $\bar{t}_c$  increases with  $N_e$ , although not strongly so, scaling as  $\sim N_e^{0.2}$  under most conditions when  $c < s$  (Fig. 2). Again, recombination rates in the neighborhood of  $s/2$  can magnify the rate of adaptation up to severalfold, with the quantitative effect being diminished when  $\delta$



**Fig. 1.** Relationship between the mean times to establishment of an adaptive two-step allele with neutral intermediate states, with the mutation rate scaling negatively with the effective population size as defined in the text (and with the resultant values of  $N_e u$  given along the upper horizontal axis). Results are given for a selective advantage of  $s = 0.01$  (solid lines and circles) and  $s = 0.0001$  (dotted lines and inverted triangles). For each  $s$ , the upper and lower solid lines, associated with sets of solid and open data points, respectively, are the results obtained by using Eq. 1 and Eq. 2 with  $c = 0$  and  $c = s/2$  (the optimal recombination rate), whereas the lines connecting open data points directly map the results for free recombination ( $c = 0.5$ ). The data were obtained as averages of 100–400 stochastic simulations of a Wright–Fisher population starting with a situation in which the zero-state **ab** is fixed.

is large. Also, as anticipated from the preceding theory, recombination rates in excess of the selective advantage of the final adaptation present a powerful barrier to establishment of the **AB** combination unless the effective population size is  $< 1/\delta$ .

The analysis is necessarily more complicated when more than two alterations are required to achieve the final adaptive state, although informative theoretical results can be obtained for a few limiting cases with neutral intermediates. For highly recombining populations ( $c > s$ ) large enough to be treated in a deterministic fashion ( $4N_e u \gg 1$ ), the time to establishment of an adaptive allele via  $d - 1$  intermediate states can be shown to be

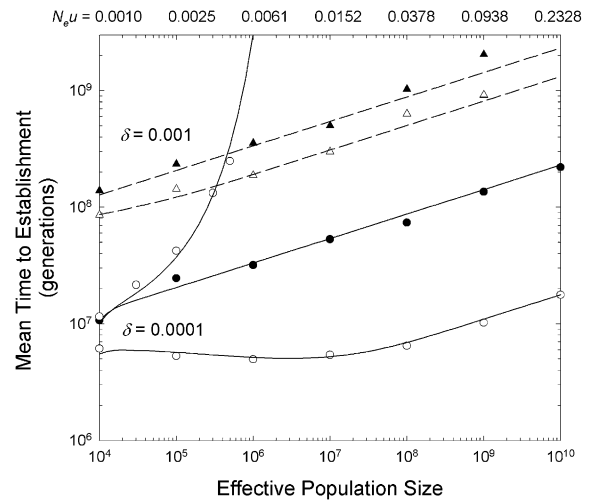
$$\bar{t}_e \approx \frac{d}{u(d-1)} (u/s)^{1/d} \quad [8a]$$

generations, which contrasts with the approximate result for complete linkage,

$$\bar{t}_e \approx \frac{1}{u(4N_e s)^{1/d}} \quad [8b]$$

(26). Here, the ratio of times to establishment with and without recombination is  $[d/(d-1)](4N_e u)^{1/d}$ , which for  $N_e u = 1$  is  $\leq 1.5$  for  $d > 2$ . Thus, the inhibitory effect of recombination on the acquisition of an adaptive allele in large populations is expected to be small when more than two intermediate neutral steps are involved, and in fact, within the range of  $N_e u$  estimates from empirical data, there is no domain of inhibition (Fig. 3).

With  $d > 2$ , one can also anticipate a reduction in the optimal rate of recombination. For example, with  $d = 3$ , the rate of creation of novel **ABC** alleles remains proportional to  $c$ , whereas the net selective advantage of such alleles is reduced to  $s - 2c$  (because there are two potential regions of recombination between the three selected sites). This result leads to an optimal  $c = s/3$ , which is in rough accord with the observations from simulated populations (Fig. 3).



**Fig. 2.** Relationship between the mean times to establishment of an adaptive two-step allele with selective advantage  $s = 0.01$  when the intermediate-state alleles have a selective disadvantage  $\delta$ , with the mutation rate scaling negatively with the effective population size as defined in the text. Results are given for  $\delta = 0.0001$  (solid lines) and  $\delta = 0.001$  (dashed lines). Lines mapping the solid points were obtained by using equations 3b, 4a, and 4b from ref. 26, which assume no recombination. The strongly upwardly bowed curve for  $\delta = 0.0001$  denotes the theoretical results for freely recombining sites, obtained with Eq. 6; this curve is off the graph scale for  $\delta = 0.001$ . For each  $\delta$ , the lower sets of data refer to the situation in which recombination is expected to maximally benefit the time to establishment ( $c = s/2$ ), with the theoretical results (curved lines) being obtained from Eq. 4b and Eqs. 5, with  $k = 100$  when  $\delta = 0.0001$  and  $k = 1,000$  when  $\delta = 0.001$ . Data points denote averages obtained from 50–400 stochastic simulations of a Wright–Fisher population starting with a situation in which the derived alleles (**A** and **B**) had zero frequencies.

Although the response of  $\bar{t}_e$  to  $N_e$  when  $d > 2$  is not as flat as in the case of the two-site model, it is still quite shallow. For example, with two neutral intermediate steps ( $d = 3$ ), there is only a 10-fold increase in  $\bar{t}_e$  over a six order-of-magnitude gradient of  $N_e$ , with the response becoming progressively flatter at large  $N_e$ , especially when the recombination rate is near optimal (Fig. 3). Even when  $d = 4$  or 5,  $\bar{t}_e$  scales with less than the square root of  $N_e$  in the absence of recombination (Fig. 3).

**Conversion to the Absolute Timescale.** Before proceeding with a summary of the implications of the theoretical results, one final scaling issue needs to be addressed. The predicted mean times of establishment of complex adaptations have been given above in units of generations, whereas generation length ( $\gamma$ ) varies dramatically among organisms. To determine the scaling of various forms of adaptive potential on an absolute timescale, such variation must be taken into account. Drawing from observations outlined in *SI Text*, it is assumed below that  $\gamma$  is proportional to  $N_e^{-0.8}$ . This assumption means that the scalings for  $\bar{t}_e$  given in the preceding section must be multiplied by  $N_e^{-0.8}$  to convert to absolute time, although the exponent in this conversion could be off by up to 20% in either direction.

As an example of the implications of this generation-length scaling for  $N_e$ -related patterns, consider the classical situation with single-site modifications involving no epistatic effects. Each generation,  $2N_e u$  mutations arise at the site, each of which has a fixation probability of  $\sim 2s(N_e/N)$  provided  $4N_e s > 1$ . The rate of adaptation is then  $4N_e u s$  per generation, which with  $u$  scaling as  $N_e^{-0.6}$  further implies a rate scaling with  $N_e^{0.4}$ . Taking the reciprocal of this quantity and multiplying by  $N_e^{-0.8}$ , the mean absolute time to establishment of a one-site adaptation then scales as  $\sim N_e^{-1.2}$ .

Now consider the situation in which the intermediate steps to a two-site adaptation are neutral. Recall from Fig. 1 that for



Now consider a two-site adaptation with a strong selective advantage of  $s = 0.02$ , in which case the optimal recombination rate for evolutionary progress is 0.01 under the two-site model. For per-site recombination rates of  $10^{-9}$ ,  $10^{-8}$ ,  $10^{-7}$ , and  $10^{-6}$ , the optimal  $c$  will then be achieved for sites separated by physical distances of  $10^7$ ,  $10^6$ ,  $10^5$ , and  $10^4$  nucleotides, respectively. These critical distances will be 10-fold smaller for adaptive combinations with  $s = 0.002$ . Because of the inclusion of introns, many pairs of coding nucleotides in the genes of multicellular eukaryotes are separated by as many as  $10^4$ – $10^5$  nucleotides, more so in the bloated genomes of vertebrates and land plants. Consequently, the recombinational environment of such species appears to be adequate to sometimes elevate the rate of procurement of complex adaptations involving sufficiently spaced sites at the within-gene level. On the other hand, the rate of acquisition of adaptations involving the joint changes at pairs of noncontiguous genes will generally be affected only minimally by recombination unless  $c > s$  (genes on different chromosomes), in which case recombination will strongly impede the rate of adaptation if the intermediate states are deleterious. In contrast, because prokaryotic genomes generally contain  $<10^7$  nucleotide sites, we can expect almost all complex adaptations with  $>1\%$  selective advantage (involving both within- and among-locus changes) to be procured at rates similar to those predicted in the absence of recombination.

In summary, the preceding results suggest that some general scaling properties may exist for the rapidity with which various

types of adaptations can be assimilated in different population-genetic contexts. In particular, prokaryotes appear to be much more efficient than eukaryotes at promoting simple to moderately complex molecular adaptations, and substantially so for those involving joint changes at different genetic loci. In contrast, adaptations requiring three or more novel mutations may arise more frequently in small populations, regardless of the level of recombination between selected sites. In the absence of comprehensive information on the molecular basis of adaptation in multiple lineages (i.e., the typical number of sites involved and their degree of epistatic interactions), these general predictions are currently difficult to test. Nevertheless, the ideas presented herein are likely to bear significantly on a number of ongoing controversies regarding the nature of adaptation, including the barriers imposed by adaptive valleys in a fitness landscape (22, 40), the role of compensatory mutation in evolution (41), and the relative rates of incorporation of adaptive and nonadaptive mutations in various lineages (42–44).

**ACKNOWLEDGMENTS.** I am very grateful to M. Hahn, R. Neher, A. Orr, B. Shraiman, and D. Weinreich for critical comments and to the Kavli Institute of Theoretical Physics for supporting my short visit during which this project was completed. This work was supported by National Institutes of Health Grant R01 GM036827 and National Science Foundation Grant EF-0827411 (to M.L.).

- Lynch M (2006) The origins of eukaryotic gene structure. *Mol Biol Evol* 23:450–468.
- Lynch M (2007) *The Origins of Genome Architecture* (Sinauer, Sunderland, MA).
- Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26:345–352.
- Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10:195–205.
- Lynch M (2008) The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180:933–943.
- Smith JM (1970) Natural selection and the concept of a protein space. *Nature* 225:563–564.
- Gillespie JH (1983) A simple stochastic gene substitution model. *Theor Popul Biol* 23:202–215.
- Orr HA (2000) Adaptation and the cost of complexity. *Evolution* 54:13–20.
- Orr HA (2002) The population genetics of adaptation: The adaptation of DNA sequences. *Evolution* 56:1317–1330.
- Rokyta DR, Beisel CJ, Joyce P (2006) Properties of adaptive walks on uncorrelated landscapes under strong selection and weak mutation. *J Theor Biol* 243:114–120.
- Martin G, Lenormand T (2008) The distribution of beneficial and fixed mutation fitness effects close to an optimum. *Genetics* 179:907–916.
- Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176:1759–1798.
- Neher RA, Shraiman BI, Fisher DS (2010) Rate of adaptation in large sexual populations. *Genetics* 184:467–481.
- Ffrench-Constant RH, Daborn PJ, Le Goff G (2004) The genetics and genomics of insecticide resistance. *Trends Genet* 20:163–170.
- Hubbard JK, Uy JA, Hauber ME, Hoekstra HE, Safran RJ (2010) Vertebrate pigmentation: From underlying genes to adaptive function. *Trends Genet* 26:231–239.
- Chan YF, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327:302–305.
- Gillespie JH (1984) Molecular evolution over the mutational landscape. *Evolution* 38:1116–1129.
- Higgs PG (1998) Compensatory neutral mutations and the evolution of RNA. *Genetica* 102–103:91–101.
- Carter AJR, Wagner GP (2002) Evolution of functionally conserved enhancers can be accelerated in large populations: A population-genetic model. *Proc Biol Sci* 269:953–960.
- Komarova NL, Sengupta A, Nowak MA (2003) Mutation-selection networks of cancer initiation: Tumor suppressor genes and chromosomal instability. *J Theor Biol* 223:433–450.
- Iwasa Y, Michor F, Nowak MA (2004) Stochastic tunnels in evolutionary dynamics. *Genetics* 166:1571–1579.
- Weinreich DM, Chao L (2005) Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* 59:1175–1182.
- Durrett R, Schmidt D (2008) Waiting for two mutations: With applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics* 180:1501–1509.
- Schweinsberg J (2008) The waiting time for  $m$  mutations. *Electron J Probab* 13:1442–1478.
- Weissman DB, Desai MM, Fisher DS, Feldman MW (2009) The rate at which asexual populations cross fitness valleys. *Theor Popul Biol* 75:286–300.
- Lynch M, Abegg A (2010) The rate of establishment of complex adaptations. *Mol Biol Evol* 27:1404–1414.
- Fisher RA (1930) *The Genetical Theory of Natural Selection* (Oxford Univ Press, Oxford).
- Muller HJ (1932) Some genetic aspects of sex. *Am Nat* 66:118–138.
- Moran PAP (1962) *The Statistical Processes of Evolutionary Theory* (Clarendon, Oxford).
- Christiansen FB, Otto SP, Bergman A, Feldman MW (1998) Waiting with and without recombination: The time to production of a double mutant. *Theor Popul Biol* 53:199–215.
- Neher RA, Shraiman BI (2009) Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proc Natl Acad Sci USA* 106:6866–6871.
- Iwasa Y, Michor F, Komarova NL, Nowak MA (2005) Population genetics of tumor suppressor genes. *J Theor Biol* 233:15–23.
- Eshel I, Feldman MW (1970) On the evolutionary effect of recombination. *Theor Popul Biol* 1:88–100.
- Walsh JB (1982) Rate of accumulation of reproductive isolation by chromosome rearrangements. *Am Nat* 120:510–532.
- Provine WB, ed (1986) *Evolution: Selected Papers of Sewall Wright* (Univ Chicago Press, Chicago).
- Goodnight CJ (2006) Peak shifts in large populations. *Heredity* 96:5–6.
- Barton NH, Otto SP (2005) Evolution of recombination due to random drift. *Genetics* 169:2353–2370.
- Keightley PD, Otto SP (2006) Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* 443:89–92.
- Hartfield M, Otto SP, Keightley PD (2010) The role of advantageous mutations in enhancing the evolution of a recombination modifier. *Genetics* 184:1153–1164.
- Wade MJ, Goodnight CJ (1998) The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. *Evolution* 52:1537–1553.
- Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA (2010) Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464:279–282.
- Eyre-Walker A (2006) The genomic rate of adaptive evolution. *Trends Ecol Evol* 21:569–575.
- Hahn MW (2008) Toward a selection theory of molecular evolution. *Evolution* 62:255–265.
- Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the *Drosophila* genome? *PLoS Genet* 5:e1000495.