J. theor. Biol. (2000) **207,** 495–511
doi:10.1006/jtbi.2000.2189, available online at http://www.idealibrary.com on IDEAL®

Gene Conversion may aid Adaptive Peak Shifts

THOMAS F. HANSEN*, ASHLEY J. R. CARTER AND CHI-HUA CHIU†

Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect St., P.O. Box 208106, New Haven, CT 06520, U.S.A.

(*Received on 9 March 2000, Accepted in revised form on 30 August 2000*)

Gene conversion is often viewed as a homogenizing force that opposes adaptive evolution. The objective of this study is to suggest a potential role for gene conversion in adaptive evolution of proteins through aiding the transfer of a population from one adaptive peak to another. Our hypothesis starts with the observation that a tandem gene duplication may result in an extra gene copy that is released from selective constraints. In such cases, individually deleterious mutations may accumulate on the extra copy of the gene, and through gene conversion these mutations may subsequently be presented to the functioning gene for selection *en masse*. Thus, groups of mutations that jointly confer a selective advantage may regularly be made available for selection. We present a mathematical model of this process and identify the range of rates of gene conversion, gene duplication and mutation under which it may operate. The results indicate that the process may be biologically feasible if the rate of appearance of the potentially beneficial mutations is not too small in relation to the rates of null mutation and of gene conversion. This process appears to be a possible mechanism for effecting adaptive peak shifts in large populations. We show that all the evolutionary steps in the proposed model may have occurred in the evolution of primate γ-globin genes. We suggest that hide-and-release mechanisms for genetic variation may constitute a more general principal of evolvability.

Introduction

Modern evolutionary biology has experienced spectacular successes in proposing and testing hypotheses about plausible evolutionary paths to the summit of adaptive peaks (Dawkins, 1996). However, it is clear that evolution may not always find a smooth ascent towards a single well-defined optimum (Lewontin, 1978; Gould, 1989; Dawkins, 1996). Indeed, any reasonably complex structure consisting of many interlocking parts is prone to generate a rugged fitness landscape with many adaptive peaks and valleys (Kauffman, 1993). While natural selection is the only known mechanism that consistently effects the movement of a population towards higher fitness, it has a shortcoming as an optimization algorithm on rugged fitness landscapes because it tends to find the nearest local optimum and thereafter oppose further evolutionary change. This provides a difficulty for the Darwinian task of finding an unbroken series of individually advantageous steps towards highly refined adaptations such as may be found in the elaborate structure of tens of thousands of proteins in every organism.

This problem has generated an interest among theoretical biologists in finding mechanisms that

*Present address: Department of Biological Science, Florida State University, Tallahassee, FL 32306, U.S.A.
†Author to whom correspondence should be addressed. E-mail: chi-hua.chiu@yale.edu

0022–5193/00/240495 + 17 $35.00/0

may effect the transition of a population from one local optimum to another. While some theories put forth have sidestepped the problem by denying the ruggedness of fitness landscapes or by postulating mechanisms that can smoothen or change the landscape (see discussion), random genetic drift is an essential element in all theories that actually allow a population to pass through a state of reduced fitness. The most famous example is Wright's (1931, 1932) shifting-balance theory, which in essence posits that random genetic drift alone can shift a small isolated subpopulation to the new optimum, where migrants may then induce the adaptive transfer in the population at large. However, the problem with genetic drift as a mechanism for peak shifts is that it does not work unless the adaptive valley is extremely shallow or the population size is extremely small (Lande, 1985; Barton & Rouhani, 1987). Even a population of a few hundred individuals is simply not able to pass through a state of substantially reduced fitness on an evolutionarily relevant time-scale.

In this paper, we propose a novel mechanism of some generality for how a gene may pass a fitness valley to a higher adaptive peak in large populations. The mechanism is founded on the observation that duplicated genes can undergo gene conversion. When a gene is duplicated, one of the duplicates may be released from selective constraints through redundancy or silencing. It is then free to acquire mutations that would otherwise be selected against. If the two duplicates are evolving in concert through gene conversion, at a rate that is not much in excess of the mutation rate, the functional copy of the gene will be repeatedly presented with groups of novel mutations that have accumulated on the other duplicate. In effect, the mechanism allows the population to explore a larger area of the fitness landscape by testing several mutations at once.

At least since Ohno's (1970) prescient work, gene duplication has been viewed as an important mechanism for generating novel adaptation. By duplicating a gene, or indeed any structure, one of the duplicates may be released from the selective constraints of the original function and is free to explore new possibilities. By now there is little doubt that duplication is a major source of novelty in protein evolution (Raff, 1996). Indeed, at least in multicellular organisms, we suspect that the vast majority of genes may have originated through duplication from another gene with a different function [but see Hughes (1994) and Force *et al.* (1999)]. The mechanism proposed here is motivated by the general importance and abundance of gene duplication in evolution, but differs from previous population genetic investigations of the fate of duplicated genes (e.g. Spofford, 1969; Ohta, 1987; Walsh, 1995; Nowak *et al.*, 1997), in that it is not primarily concerned with the functional diversification of duplicated loci, but rather with the role of duplication in the adaptation of a single gene.

In contrast to gene duplication, gene conversion has rarely been assigned a creative role in evolution, and is usually seen as a homogenizing force that prevents the evolutionary diversification of duplicated genes (Walsh, 1987). It has been shown that gene conversion may benefit individual fitness by homogenizing duplicated genes that carry deleterious mutations when there is negative epistasis between the mutations (Hurst & Smith, 1998). Gene conversion may similarly reduce the mutation load of a multigene family (Ohta, 1989). Concerted evolution may also aid the spread of a uniformly advantageous allele from one duplicate to others (Ohta & Dover, 1984; Slatkin, 1986; Dover, 1993; Inomata & Yamazaki, 1996; Hurst & Smith, 1998). Our proposed mechanism adds another possible link between gene conversion and adaptation. Unfortunately, the molecular mechanisms of gene conversion are still not completely elucidated, and the rates and conditions under which it occurs are poorly known (Elder & Turner, 1995). As will be seen, our hypothesis requires that the rate of gene conversion is limited to a certain range that is determined by the mutation rate, the population size and the strength of selection.

In the following, we present an analytical model of the process and use this to compute probabilities and mean times for crossing a two-step adaptive valley as functions of population size, strength of selection, rates of mutation, rate of gene duplication and rates of gene conversion. These results are then compared with the rate and mean time of peak shift through genetic drift. We also investigate the process with simulations of a more biologically detailed model which

assumes small population size. We hypothesize that the process may operate under biologically realistic parameter values and may be general to protein evolution. In the discussion, we suggest that the evolution of fetal specificity of primate γ-globins (Chiu *et al.*, 1999) is a possible empirical example of this process.

Model and Methods

As illustrated in Fig. 1, we start with a gene in state 1 and consider the task of passing through states 2 and 3 with successively lowered fitness to state 4 which is more fit than state 1. We consider a two-step adaptive valley as this allows us to see how the process scales with an additional step without making the results unduly complicated. By the term "state" we mean steps in a process and not specific alleles. The model does not assume that specific allelic states have to be traversed in order.

As illustrated below, a very small or a very large population may traverse the fitness valley through genetic drift or random occurrence of double mutations, respectively. However, our primary interest is in the more complicated path illustrated in Fig. 2. The first step is the duplication of the gene into two duplicates that are both in state 1. We denote a population as in state "11" when it is fixed for this duplicate. If one of the gene copies is silenced after the duplication, or if the expression of the duplicates is redundant and regulated by downstream factors, it is reasonable to assume that one of the gene copies can freely accumulate mutations under relaxed selective constraints. For now, we will focus on the former situation in which one duplicate is silenced; however, we later discuss adjustments to the model that take into account expression of both gene copies following the duplication.

The next step is the appearance of a mutation from state 1 to state 2 in the silenced duplicate (Fig. 2). As this gene is not expressed, this 12-haplotype may drift to fixation. When this happens, we denote the population as being in state 12. Similarly, further mutations can appear on the silent duplicate and drift to fixation, bringing the population successively into states 13 and 14. Finally, in state 14, gene conversion can effect the transition of the expressed gene from state 1 to state 4 so that it can be brought to fixation by selection. If this happens, the population is in

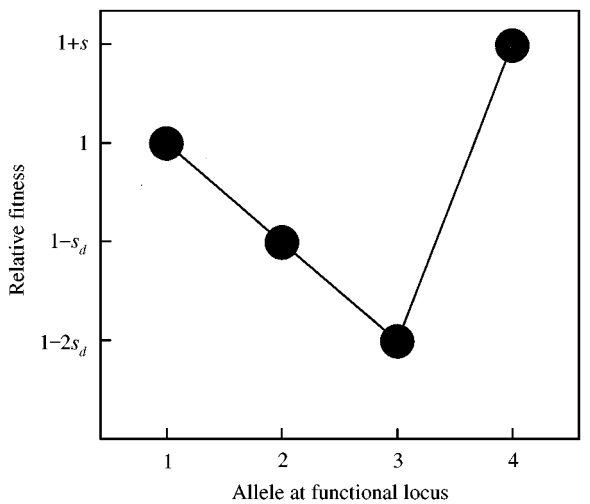

FIG. 1. Illustration of the fitness valley to be crossed. The four states of the gene are on the x-axis. The valley is to be crossed in the sequence $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$. The parameter s represents the heterozygote selective advantage of the new peak relative to the old, while the parameter s_d is the heterozygote selection coefficient against each step down in the valley.

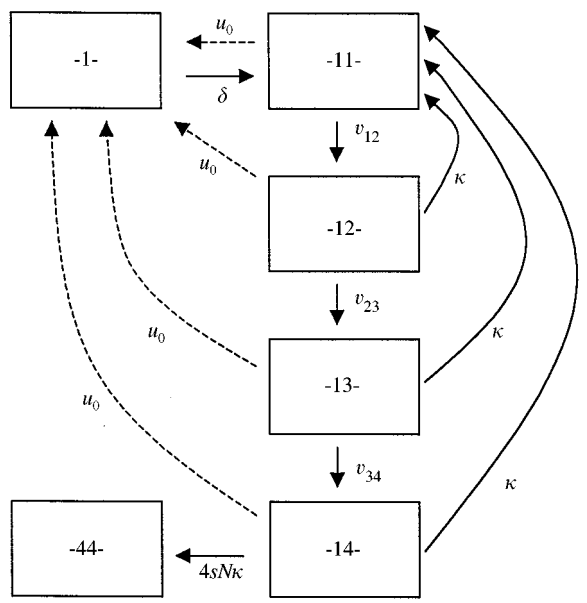

FIG. 2. Flow diagram depicting the model. The boxes represent populations being fixed for different states of the duplicated gene. The first number represents the state of the gene being expressed, while the second number represents the state of the unexpressed (or redundant) gene. The parameters along the arrows are rates of transition from one state to another. The -1-box represents the pre-duplication stage, and the adaptive peak is reached when the population ends in the -44- box. See main text for explanations.

state 44 and is considered to have reached the new adaptive peak.

However, along this path several other possibilities must be considered. First, gene conversion may occur at every step along the path, and not just in the final step. If the expressed gene is converted to state 2 or 3, the effect is deleterious and we assume that such an allele is always eliminated by selection. However, the silent duplicate can also be converted back to state 1 by the expressed gene at any point in the process. This convert is then neutral and may drift to fixation bringing the population back to state 11. If the gene conversion rate is too high, we see that this may effectively block the process. The other possibility is that the silent duplicate may acquire mutations that are not on the path to the new adaptive peak. Such mutations may block the passage to the new peak by blocking the concerted evolution of the two duplicates, or by having deleterious effects that are sufficiently severe to overshadow the advantage of state 4 when it emerges. If such mutations appear, the process has to start all over again from state 1.

Let δ denote the rate of duplication per allele per generation. In the "silent-duplicate" model this only includes cases where the duplicate is truly silenced after the duplication. The rate of substitution of the duplicate is then $2N\delta 1/2N = \delta$, where N is the population size. Let κ denote the rate of gene conversion per duplicate per generation. We assume this is constant and unbiased throughout the process. When the silenced duplicate is converted, the product is neutral, and the rate of substitution is κ. However, at the last step of the process, from 14 to 44, the rate of substitution is approximately $2N\kappa 2s = 4sN\kappa$, where s is the heterozygote coefficient of selection in favor of state 4 over state 1 (minus the effect of any deleterious mutations that may have accumulated on the duplicate throughout the process). The approximation $2s$ for the fixation probability of a new advantageous allele is good provided sN is not smaller than one and s is not too large.

Let v_{12}, v_{23}, and v_{34} be the rates of mutation per allele per generation for the transitions $1 \to 2$, $2 \to 3$ and $3 \to 4$, respectively. As these mutations are neutral in the silenced locus, the rate of substitution equals the mutation rate. Note that these mutation rates may be smaller than typical estimates of locus-specific mutation rates, as there only will be a small subset of possible mutations that actually are a part of the path to the new optimum. It is reasonable to assume that $v_{12} > v_{23} > v_{34}$, as fixation of the first mutation may narrow the target for further changes, but as it will turn out that the v's can be interchanged without affecting the results, we will without loss of generality assume that they all have the same value, v. Let u_0 be the rate of mutations per allele per generation that are either able to block the concerted evolution of the two duplicates or are so deleterious when expressed that they can outweigh the fitness benefit of the new adaptive peak. We will refer to u_0 as the null-mutation rate. When appearing on the silent duplicate, the rate of substitution of these null mutations is u_0. If the accumulation of many mutations with minor deleterious effects is the main source of decay of the silent duplicate, it might be reasonable to increase the probability of "losing" the duplicate as time progresses. However, if insertions, deletions or other mutations with major effects are the main risk, then it is more accurate to keep the null-mutation rate constant throughout the process. Mainly for simplicity, we keep u_0 constant.

One caveat is that the analysis of this model assumes that all fixations happen instantaneously in time. We do not consider the possibility that a second mutation or gene conversion happens when another is segregating. In a very large population this assumption will be invalid.

The mathematical description of the model and the necessary calculations for producing our results are given in the appendices. In the results we present two statistics to assess the feasibility of peak shifts. The first is the probability, α, of effecting a peak shift given that a duplication has occurred and become fixed in the population. The actual rate of peak shifts in an ensemble of genes or populations challenged with same adaptive valley is then $\lambda = \alpha\delta$. This is computed in Appendix A. The second statistic, computed in Appendix B, is the expected waiting time, $\bar{t}$, for a peak shift to occur starting from the pre-duplication stage. For comparison, we present a similar analyses of peak shift through pure genetic drift and fixation of double mutants in Appendices C and D.

In evaluating the results, it is crucial to decide what constitutes an evolutionarily realistic rate of peak shift. This is somewhat subjective and the reader may have opinions different from ours. We will judge rates of peak shift up until the order of 10^{-8} per generations and waiting times up to the order of 10^{8} generations to be evolutionarily relevant. It does not seem unrealistic that a gene or a protein may sit trapped on a local adaptive peak for 100 million generations. This is especially likely if the improvement relates to internal biochemical function, which may be buffered against changes in the external environment. Furthermore, from the perspective of a genome with tens of thousands of genes, rates on this order may still make the process regular in the genome as a whole. Finally, when comparing the feasibility of the gene-conversion model with the shifting-balance model, one should remember that the latter postulates a large number of subpopulations in which a peak shift can occur. Doing the same for the gene-conversion model may increase the rate of peak shift with a factor equal to the number of subpopulations (except when population size is the limiting factor).

To assess the accuracy of the theoretical approximations, we also computed mean time to peak shift from a set of individual-based stimulations as described in Appendix E. The simulations showed that the analytical results were very accurate over a range of parameter values with N ranging from 100 to 1000. Hence, we do not report detailed results from the simulations.

Results

PEAK SHIFT IN THE GENE-CONVERSION MODEL

The full formulas for the probability and expected time to peak shift are given in Appendices A and B. Here we illustrate their main properties with several approximations based on restricting the range of parameter values. A set of numerical examples is given in Table 1.

Both the rate of peak shift and the inverse of the expected time are increasing monotonically with sN and rapidly reach an asymptote when u_0 becomes insignificant in comparison with $4sN\kappa$. Hence, unless the gene-conversion rate is much smaller than the null-mutation rate, the process is insensitive to sN. Unless otherwise

TABLE 1

Probabilities and times to peak shift under the gene-conversion model

v	u_0	κ	δ	sN	α	$\bar{t}$
10^{-8}	10^{-6}	10^{-6}	10^{-6}	∞	2.5×10^{-7}	8.1×10^{12}
10^{-7}	10^{-6}	10^{-6}	10^{-6}	∞	0.00022	9.3×10^{9}
10^{-6}	10^{-6}	10^{-6}	10^{-6}	∞	0.071	2.7×10^{7}
10^{-6}	10^{-7}	10^{-7}	10^{-6}	∞	0.73	5.0×10^{6}
10^{-5}	10^{-6}	10^{-6}	10^{-6}	∞	0.73	1.7×10^{6}
10^{-7}	10^{-6}	10^{-6}	10^{-6}	5	0.00020	1.0×10^{10}
10^{-6}	10^{-6}	10^{-6}	10^{-6}	5	0.065	3.0×10^{7}
10^{-6}	10^{-6}	10^{-5}	10^{-6}	∞	0.0063	3.2×10^{8}
10^{-6}	10^{-8}	10^{-5}	10^{-6}	∞	0.43	1.4×10^{8}
10^{-6}	10^{-6}	10^{-6}	10^{-5}	∞	0.071	1.4×10^{7}
10^{-6}	10^{-6}	10^{-6}	10^{-7}	∞	0.071	1.5×10^{8}

Note: The last two columns show the probability of peak shift given that a duplication has occurred, α, and the expected time it takes before the population arrives at the new peak from the pre-duplication stage, $\bar{t}$, as functions of parameter values. The parameters are the forward-mutation rate, v, the null-mutation rate u_0, the gene-conversion rate, κ, the gene-duplication rate, δ, the population size, N, and the selective advantage of the new peak over the old, s. The computation of α and $\bar{t}$ are based on the exact eqns (A.8) and (B.2). The first five rows show the effect of increasing v relative to the other parameters. The next two rows illustrate that the process is relatively insensitive to changes in populations size and the selective advantage of the new peak. The next two show some effects of increasing κ relative to u_0, and the last two rows show some effects of altering the gene duplication rate.

mentioned, we will therefore assume that sN is infinity.

Case 1: *All rates equal.* Here we assume that $v = u_0 = \kappa = \delta$. Then, expressing the results in terms of δ, the rate of peak shift becomes $\lambda = \delta/11$ and the expected time becomes $\bar{t} = 27/\delta$. Thus, peak shifts are about one order of magnitude slower and more rare than gene duplications. We judge the process as biologically realistic in this case unless $\delta = v = u_0 = \kappa$ is smaller than about 10^{-7}.

Case 2: *Forward-mutation rates smaller than null-mutation rates.* In this case, we make approximations based on $v \ll u_0$:

$$\alpha = \frac{v^3}{u_0(u_0+\kappa)^2}, \qquad \bar{t} = \frac{(u_0+\delta)(u_0+\kappa)^2}{v^3\delta}. \quad (1)$$

In addition to the rate of duplication, the major limitations on the rate of peak shift is either the ratio $(v/u_0)^3$, or, if the gene-conversion rate is larger than the null-mutation rate, the ratio $v^3/u_0\kappa^2$. If we assume κ smaller than u_0 and that v is one order of magnitude below u_0, the rate becomes $\lambda = 10^{-3}\delta$, or one-thousandth of the gene-duplication rate. If rate of gene duplication are as high as 10^{-5} this falls into our range of evolutionary realistic rates. However, even when δ is very high, the expected time is $\bar{t} = u_0^2/v^3$, which demands rates of forward mutations on the order of 10^{-5} to give an expected time as low as 10^8 generations. In conclusion, the process is unrealistic if the forward-mutation rate is an order of magnitude or more less than the null-mutation rate. Note though that the process speeds up an order of magnitude if it takes only one mutational step, instead of two, to cross the adaptive valley.

Case 3: *Null-mutation rates small.* If we assume that u_0 is much smaller than κ and δ we get

$$\alpha = \frac{v^3}{u_0(\kappa^2 + 3v(\kappa+v)) + v^3},$$

$$\bar{t} = \frac{\delta(\kappa^2 + 3v(\kappa+v)) + v^3}{v^3\delta}. \quad (2)$$

Note that the probability of peak shift approaches unity as the null-mutation rate goes to zero. However, this does not necessarily mean that the process is biologically relevant as it still may take a forbiddingly long time to arrive at the new peak if gene conversion is too frequent. If the duplication and conversion rates are smaller than the rates of forward mutation, the expected time approaches $1/\delta$ and the process is limited by the waiting time for a duplication to occur. Hence, when v is equal to or larger than the other parameters, the process is biologically realistic.

Case 4: *Gene-conversion rates large.* If the rate of gene conversion is too large relative to the rates of forward mutation, the process may never be able to reach the final step. If we assume that κ is much larger than u_0 and v (and that $u_0\kappa^2 \gg v^3$), we get

$$\alpha = \frac{v^3}{u_0\kappa^2}, \qquad \bar{t} = \frac{(u_0+\delta)\kappa^2}{v^3\delta}. \quad (3)$$

Hence, a rate of gene conversion two orders of magnitude above the rates of forward mutations may produce unrealistic rates and times of peak shifts. However, if the rate of gene conversion is only one order of magnitude above the forward-mutation rates, peak shifts may still be realistic. Assume, for example, that $v = u_0 = \delta = 10^{-6}$ and $\kappa = 10^{-5}$, then we get $\alpha = 0.01$ and $\bar{t} = 2 \times 10^8$, which may be biologically realistic.

Case 5: *Gene-conversion rates small.* If the rate of gene conversion is much smaller than the rate of null mutation, the outcome depends strongly on sN. If we assume that $4sN\kappa$ is still much larger than u_0, it can be shown that the process depends mainly on the ratio of forward to null-mutation rates and behaves essentially as in cases 1 and 2. If, on the other hand, $4sN\kappa$ is much smaller than u_0, we get

$$\alpha = \frac{4sN\kappa v^3}{u_0^2(u_0^2 + 3v(u_0+3v)) + u_0v^3},$$

$$\bar{t} = \frac{u_0(u_0+\delta)(u_0^2 + 3v(u_0+v)) + u_0v^3}{4sN\kappa v^3\delta}. \quad (4)$$

The process is now limited mainly by the rate of gene conversion through $4sN\kappa$. If, for example, $v = u_0$, we get $\alpha = sN\kappa/2v$, and, if in addition $\delta \gg v$, we get $\bar{t} = 5/(4sN\kappa)$. Hence, with rates of forward and null mutations equal, $4sN\kappa$ need only be of the order of about 10^{-8} to make the process biologically realistic. This means that the process is feasible in a large population even when gene conversion is extremely rare. Essentially, this is due to the fact that gene conversion is needed only in the last step, and in a large population there are many individuals in which the crucial, and now selectively advantageous, gene conversion may appear.

COMPARISON WITH ADVANTAGEOUS AND NEUTRAL MUTATIONS

To provide a baseline for evaluating these results, it is informative to compare with the expected time for evolution to produce an adaptation through individually advantageous steps. The expected waiting time for the fixation of a single mutation occurring at rate v with selection coefficient s is $(4sNv)^{-1}$. The successive fixation of two such mutations is $(2sNv)^{-1}$, and for three mutations, which is comparable in complexity to crossing out two-mutation valley, it is $3(4sNv)^{-1}$. In a very large population the production of such an adaptation is a rapid event, but if $sN = 1$ it takes about 7.5 million generations for $v = 10^{-7}$. The expected time to successively fix three neutral mutations by genetic drift (ignoring back mutation and the time the mutations spend segregating) is $3v^{-1}$, or, if $v = 10^{-7}$, 30 million generations, regardless of population size.

COMPARISONS WITH PEAK SHIFTS DUE TO RANDOM GENETIC DRIFT

In this section, we consider the alternative possibility of crossing the same adaptive valley through fixation of deleterious mutations. In Appendix C, we present and analyse a simple model of this situation. As before we start the process in state 1, the gene may then pass into states 2 and 3 successively by fixation of deleterious mutations. From state 3 it can go to state 4 through the fixation of a single advantageous mutation. However, we also include the possibility that the population may revert from state 3 to state 2, and from state 2 to state 1, through the fixation of advantageous back mutations. We assume these back mutations to occur with the same rate as the forward mutations and with a selective advantage equal to the selective disadvantage of the forward mutations.

Let v_{ij} denote the rate of mutation from state i to state j per locus per generation. We assume all these rates are equal and denote them collectively by v. The rate of fixation of a deleterious forward mutations is then $2NvP_d(s, N)$, where

$$P_d(s, N) = \frac{e^{2s/(1-s)} - 1}{e^{4sN/(1-s)} - 1} \tag{5}$$

is the approximate probability of fixation of a deleterious additive allele where s is the selection coefficient against the heterozygote (Bürger & Ewens, 1995). For simplicity, we assume that the selection coefficients pertaining to the transition from state 1 to state 2 is the same as those pertaining to the transition from state 2 to state 3. We denote this selection coefficient s_d to distinguish it from the selection coefficient pertaining to the transition from state 3 to state 4, which we denote s.

Under these assumptions, the rate of peak shift is approximately

$$\lambda_d = \frac{2sNvP_d(s_d, N)^2}{2s_d(s + s_d) + sP_d(s_d, N)} \tag{6}$$

and the expected time to arrive at the new peak is

$$\bar{t}_d = \frac{4s_d(s + s_d) + 2(s_d + 2s)P_d(s_d, N) + P_d(s_d, N)^2}{4sNP_d(s_d, N)^2 v} \tag{7}$$

and assuming that $P_d(s_d, N)$ is much smaller than s_d and s, it becomes

$$\bar{t}_d = \frac{s_d(s + s_d)}{sNP_d(s_d, N)^2 v}. \tag{8}$$

Note that the inverse of the rate becomes identical to the expected time as the deleterious fixation probability becomes much smaller than the selection coefficients.

In Table 2, we give some numerical examples of the time and probability of peak shifts due to drift. The table reveals that a two-step adaptive valley cannot practically be crossed by genetic drift if s_dN is larger than 1. However, if s_dN is less than 1, genetic drift will usually be more likely than gene conversion as a mechanism for peak shifts.

In very large populations peak shifts may also occur through the random occurrence and fixation of double mutants (Gillespie, 1984; Phillips, 1996). In Appendix D, we consider this possibility. We derive the stochastic equilibrium distribution of double deleterious mutants in a large population and use this to derive the rate of peak shift due to the occurrence of the final advantageous mutation in a segregating double mutant. Under this model the expected rate of peak shift is

$$\lambda_{dm} = \frac{4sNv^3}{s_{d1}s_{d2}}, \tag{9}$$

where s_{d1} and s_{d2} are the selection coefficients against the single and the double deleterious states, respectively, and s is the selection coefficient in favor of the advantageous state when it appears. The expected time to peak shift is the inverse of λ_{dm}. Assume for illustration that $v = 10^{-7}, s_{d1} = 0.01, s_{d2} = 0.02$ and $s = 0.1$. Then N must exceed 10^{10} before λ_{dm} reaches 10^{-8}. With v of the order of 10^{-6}, the population size must exceed 10^7. Clearly, this mechanism of peak shift is an alternative only in extremely large populations. Note though that λ_{dm} increases with a factor s_{d2}/v when the valley contains only one deleterious step. In the above examples, this would reduce the necessary population sizes with 4–5 orders of magnitude and make the mechanism much more general.

TABLE 2
Probabilities and times to peak shifts under the genetic-drift model

v	N	s_d	s	α	$\bar{t}$
10^{-6}	100	0.01	0.03	0.0010	1.0×10^{9}
10^{-6}	300	0.01	0.03	2.8×10^{-10}	3.6×10^{15}
10^{-6}	500	0.01	0.03	4.4×10^{-17}	2.3×10^{22}
10^{-6}	100	0.05	0.15	9.6×10^{-18}	1.0×10^{23}
10^{-5}	1000	0.001	0.20	0.0014	7.5×10^{7}

Note: The last two columns show the probability, α, of a peak shift given that the first mutation occurs and the expected time, $\bar{t}$, it takes to arrive at the new peak for a population where the first mutation has not yet occurred. The parameters are explained in the main text.

INTERPRETATIONS AND EXTENSIONS OF THE GENE-CONVERSION MODEL

So far we have been working under the assumption that immediately following the gene duplication, one of the duplicate copies is completely silent and therefore released from selective constraints. However, this is not the only outcome of a gene duplication event, and may not even be the most frequent. Thus, we will briefly consider some alternatives and then discuss the implications to our proposed model.

One possibility is that both duplicates are similarly expressed. The consequences of this depend on how the genes are regulated. In some cases, the total rate of production of a biologically relevant product of gene expression may change, a situation that may lead to reduced fitness and therefore preclude the fixation of the duplication in the first place. In a few cases altered expression may lead to increased fitness and the duplicate may fix in the population, but it will not be suited to aid a peak shift as mutations on either duplicate will presumably reduce fitness. However, if the biologically relevant products downstream of these genes are regulated to a specific level that is insensitive to the genes' rate of expression, the genes will be redundant with respect to fitness. This situation is essentially captured by our model, as the duplicate may drift to fixation and mutations accumulate on one (but not both) of the gene copies without having deleterious effects. In this case, the only change to the model would be to double the mutation rates at the 11 state, as either of the duplicates could start accumulating mutations. After this step, the model would be the same and our qualitative conclusions remain unchanged.

It is not known how commonly a duplication may lead to complete silencing or redundancy of a gene copy. It is possibly more common for a duplicate to have some slight residual effects on fitness resulting from a low, but not ignorable, level of expression of the copy, slight alterations

of the regulatory equilibrium in the case of redundancy, or side effects related to competition for metabolites or transcription factors. As long as the population is not very large, the outcome may still aid peak shifts, as slightly deleterious mutations may occasionally drift to fixation when $s_d N < 1$. In fact, slight selection on the duplicate may aid the peak shift by reducing the rate of accumulation of null-mutations much more than it hinders the accumulation of forward mutations. In essence, its effect is equivalent to reducing v with one factor and reducing u_0 by a larger factor. These factors will increase with population size. If the rate of null mutation is large such changes may speed up the rate of peak shift considerably.

In general, selective advantages or disadvantages of the gene duplication, gene conversion or forward-mutation steps may be incorporated into the model simply by reinterpreting the parameters in terms of the appropriate substitution rates. For example, it may not be uncommon for a gene duplication to confer a selective advantage, and Clark (1994) has shown that recurrent duplication of a gene may reduce its mutation load and propel the duplicates to high frequency. Ohta (1989) has similarly shown that gene conversion may reduce the mutation load of a gene family and thus aid the invasion of a duplication. Thus, it is possible that the rate at which duplications are incorporated into the population is much higher than the mere haploid duplication rate, δ.

Discussion

The purpose of this paper was to put forward and investigate the hypothesis that gene conversion may aid the adaptive evolution of proteins by presenting groups of mutations to selection simultaneously. If the mutations interact in their effects on fitness this may facilitate adaptation. In the extreme case, mutations that are individually deleterious, but advantageous as a group, may be fixed by selection if they can accumulate on a duplicated gene with relaxed selective constraints and then be simultaneously presented for selection through a gene conversion. Our modeling has identified a set of evolutionary parameter values under which such a process may operate.

The most critical difficulty with this hypothesis is that the rate of accumulation of potentially beneficial mutations on the duplicate cannot be very much smaller than the rates of null mutation and of gene conversion. Indeed, if the forward-mutation rate is one order of magnitude lower than the null-mutation rate, the crossing of a two-step adaptive valley already seems unlikely. In assessing this result it is important to remember that the null-mutation rate only includes mutations with large effects, including deletions or insertions, that seriously impair functionality or block gene conversion. Weakly deleterious mutations may freely accumulate up to the point where their cumulative effects outweigh the selective advantage of the new adaptive peak. In any case, the hypothesis probably requires that forward mutation rates are on the order of 10^{-6} or more. This means that each forward step must be achievable through many possible nucleotide substitutions.

There seems to be a considerable range of gene-conversion rates that may allow peak shifts. As a rule of thumb, a two-step peak shift may occur when the parameter $4sN\kappa$ is larger than about 10^{-8} and the rate of gene conversion itself is not more than an order of magnitude in excess of the forward mutation rate. Gene-conversion rates have been studied in mammalian cell lines that contain inserted sequences in which conversion events produce observable phenotypes. Estimated per cell-generation rates of gene conversion in mice and rats range from 6.6×10^{-7} to 5×10^{-5} (Liskay & Stanchelek, 1983; Rubinitz & Subrami, 1986; St Onge *et al*., 1993). Liskay *et al*. (1987) found that the rate increased with the length of the homologous sequences from 6.5×10^{-9} in 200 bp homologues to 2×10^{-6} in 1800 bp homologues. Studies like these have resulted in the use of gene conversion rates in the range of 10^{-7}–10^{-5} in theoretical studies (Walsh, 1987; Ohta, 1998). As per organism-generation rates are presumably higher than per cell-generation rates, rates of gene conversion may sometimes be too large for the process to operate. However, given the range of the estimates, it may not be uncommon to find gene-conversion rates in the appropriate interval.

The rate of gene duplication may also become the rate-limiting step if it is too low. However,

gene duplication seems to be a rather common evolutionary event. In *Drosophila melanogaster* the *rosy* locus has been estimated to duplicate at rates of 1.6×10^{-5} to 1.7×10^{-4} per generation (Gelbart & Chovnik, 1979; Shapira & Finnerty, 1986), and the *maroon-like* locus at a rate of 2.7×10^{-6} per generation (Shapira & Finnerty, 1986).

An important result is that the rate of peak shift by gene conversion increases with population size. This sets gene conversion apart from most other suggested mechanisms that rely on genetic drift during small population size to effect the peak shift. We have shown that a two-step adaptive valley is not likely to be crossed through genetic drift unless $s_d N$ is less than one. With a selective disadvantage of 1% this restricts the population size to be less than a hundred individuals. Clearly, it takes special circumstances, such as those postulated by the shifting-balance theory, to maintain population sizes this low over an evolutionary time-scale. Given that serious doubts have been raised about the shifting-balance theory (Coyne *et al.*, 1997), the mechanism suggested in this paper seems a viable alternative for protein evolution. At the other end of the spectrum, the random occurrence of double mutants may be a more realistic mechanism of peak shift in extremely large populations (Phillips, 1996). But gene conversion or similar mechanisms (see below) may be the only realistic way of effecting a peak shift for a large range of population sizes.

Other proposed mechanisms for peak shifts rely on changing the adaptive landscape. For example, Price *et al.* (1993) describe how a peak shift may result from a correlated response to selection and sexual selection may have a similar effects (Lande & Kirkpatrick, 1988; Sætre, 2000). Wagner *et al.* (1994) discuss how certain epistatic interactions may make a peak shift more feasible and Whitlock (1995) discusses how an increase in phenotypic variance during a bottleneck may smoothen the adaptive landscape. None of these mechanisms seem to have much relevance for the adaptation of single proteins. The most relevant discussion of peak shift from this perspective is provided by Gordon (1994), who considers how the increase in dimensionality of the adaptive landscape resulting from a gene duplication may allow a population to escape a local optimum.

In conclusion, we have shown that gene conversion may play a role in the adaptive evolution of proteins for a range of biologial parameter values that may or may not be realistic. However, better estimates of crucial parameters are needed to reach confidence in the mechanism. To the extent that peak shifts occur at all in protein evolution, gene conversion seems a more likely mechanism than any explicit alternative.

To illustrate our hypothesis, we now consider the evolutionary transition from embryonic to fetal expression of γ-globin genes in simian primates (Chiu *et al.*, 1999), as outlined in Fig. 3. Although we present no direct evidence for an adaptive peak shift, the example minimally illustrates the feasibility of the various genetic transitions postulated by our model.

The γ-globin gene is a member of the β-type globin family that, together with genes of the α-type globin family, encode the protein chains that form hemoglobin. In humans, the linkage order of β-type globin genes (5′-ε-γ1-γ2-$\psi\eta$-δ-β-3′) parallels the timing of expression of functional loci, with the 5′ most genes expressed earlier and the 3′ most genes expressed later in ontogeny. The expression of the β-type globin loci during development is determined in large part by the locus control region (LCR), an enhancer located far upstream of the β-globin cluster (Martin *et al.*, 1996). Phylogenetic reconstructions reveal that the tandem duplication of γ in a simian ancestor was mediated by an insertion of two LINE elements, L1a and L1b into the β-globin gene cluster, flanking the single γ gene (L1a-γ-L1b) and a subsequent unequal crossing-over event between these elements producing L1a-γ1-L1ba-γ2-L1b (Fitch *et al.*, 1991). Chiu *et al.* (1997, 1999) have proposed that immediately following the γ duplication in the simian ancestor, the newly duplicated γ2 gene may have been largely silent both during embryonic life because it was located far enough from ε and the LCR to become activated, and during fetal life because its promoter had *cis*-sequences that bind fetal repressor proteins with high affinity. In addition to base substitutions in its promoter which released binding of fetal repressors, the γ2 gene also accumulated substitutions in the coding sequences which increased the level of oxygen capture from the mother's blood, improving its function as

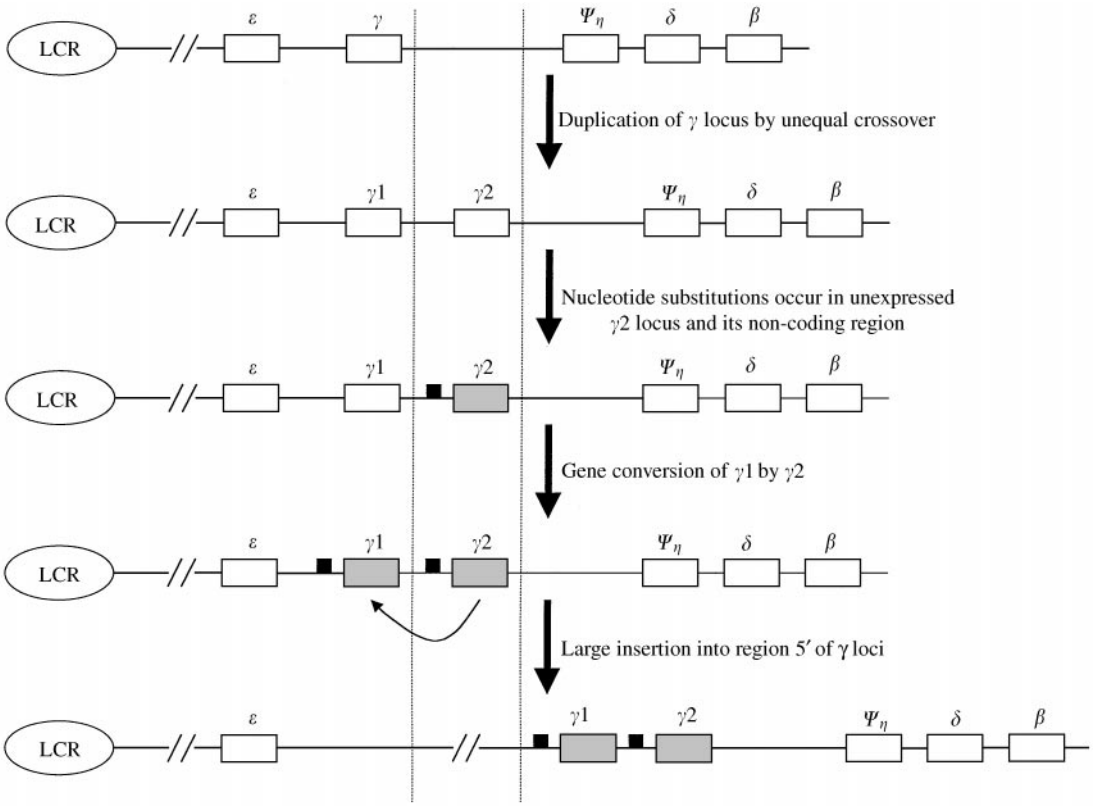

FIG. 3. Evolution from embryonic to fetal expression of γ-globin genes in simian primates. The initial duplication placed one copy of the γ-globin genes ($\gamma 2$) in a zone (between the lines) that, based on distance from the locus control region (LCR), is not active during embryonic life. The silent $\gamma 2$-gene accumulated mutations that disrupted binding of fetal repressor proteins (■) and changed amino acid residues in the coding sequence (hatching) that increased the protein's affinity for oxygen. These changes would likely have been deleterious if they happened individually in an embryonically expressed gene, but may confer an advantage in a fetally expressed gene. Gene conversion events transferred the mutations that favor fetal expression of $\gamma 2$–$\gamma 1$. This situation is most evident in extant catarrhine (Old World monkeys and hominoids) primates, where both γ genes are fetally expressed, with $\gamma 1$ expressed at a level three-fold that of $\gamma 2$. Chiu *et al.* (1999) have proposed that this catarrhine-specific pattern of γ expression resulted from a LINE element insertion between ε and $\gamma 1$ that shifted the position of the γ duplication region to a distance from the LCR that favors fetal expression of both γ genes, particularly $\gamma 1$. Simplified from Chiu *et al.* (1999).

a fetally expressed γ gene (Bunn & Forget, 1986). These changes in the $\gamma 2$ locus were then transferred to the $\gamma 1$ gene by gene conversion (Chiu *et al.*, 1997). This situation is especially evident in extant catarrhines, which express both $\gamma 1$ and $\gamma 2$ ($\gamma 1 > \gamma 2$) in fetal life (Bunn & Forget, 1986). However in platyrrhines, $\gamma 2$ is preferentially expressed and $\gamma 1$ is either non- or very weakly expressed (Johnson *et al.*, 1996; Chiu *et al.*, 1996, 1999). Although there is no direct evidence for relative fitness values of the different stages, it can be speculated that the evolution of fetal γ genes was associated with extended fetal gestational periods in simian primates.

The crucial role of gene conversion in our model is to transfer a group of jointly advantageous mutations from a silent or non-functional gene copy to a position where they can be expressed. We note that there are a variety of other molecular mechanisms that may lead to the expression of previously silent mutations. First, if the gene copy on which the mutations are accumulating is not completely silent there will be positive selection to upgrade its rate of expression (and downgrade the other copy) as soon as the beneficial group of mutations is in place, and this may happen without any change in the location of the gene. Indeed, a straightforward alternative to the last gene-conversion step is a simple regulatory mutation that brings the silent gene to expression. Another alternative is unequal recombination between the two tandem

duplicates (call them a and b) that would produce a single hybrid gene locus (locus a-b) and a triplicated product (locus a, hybrid b-a locus, locus b). After such an event it is possible that either the hybrid locus or the hybrid part of the triplicated product contain both the advantageous group of mutations from the b-locus and the parts of the a-locus necessary for proper expression. Interestingly, an unequal-crossover event between the tandemly duplicated γ loci (5′-γ1-γ2-3′) occurred independently in two different New World monkey genera (*Aotus* and *Saimiri*), producing a single hybrid 5′-γ1/γ2-3′ locus that is fetally expressed (Chiu *et al*., 1996, 1997). Whether any of these mechanisms are more likely than gene conversion to achieve the final step is an open question.

Inomata & Yamazaki (1996) discuss another intriguing candidate case where concerted evolution may be linked to adaptation in the duplicated *Amy* genes of the *Drosophila melanogaster* species subgroup. The coding regions of these genes show concerted evolution while regulatory regions are diverging. There is good evidence for positive selection being involved in both regulatory and protein evolution. The role of gene conversion is not clear, but it may very well be aiding the spread of an advantageous mutation from one duplicate to another. Dover (1992, 1993) discusses cases where gene conversion and other mechanisms of homogenization may be involved in spreading alleles over repeated elements and suggests that this may facilitate molecular coevolution with interacting genes.

Finally, we wish to point out that the mechanisms presented in this paper can be viewed as examples of a more general principle of evolvability based on the periodic release of hidden genetic variation. If the genetic system is able to hide variation for extended amounts of time, allow it to accumulate and then reexpress it *en masse* for selection, this may enhance evolvability by allowing qualitatively new variations to appear. Gene conversion is but one mechanism that can achieve this. A potentially more general mechanism resides in the robustness of the genetic architecture or the developmental system (Gerhart & Kirchner, 1997), where the effects of minor mutations are absorbed by the system until a certain threshold is reached. This is supported by the observation that mutations often have more severe effects when expressed in a novel genetic background (Moreno, 1994), and by the observation that genetic or environmental stress may act to increase genetic variation (Hoffmann & Parsons, 1997). Robustness of the genetic system may result from stabilizing selection favoring the canalization of the wild type onto a plateau where disturbances have minimal effects (Wagner *et al*., 1997). The hiding of genetic variation provided by dominance may serve as a very simple example where several mutations may potentially accumulate on a rare recessive allele before they are expressed for selection. Recently, Rutherford & Lindquist (1998) have provided a startling example of such a hide-and-release principle of evolvability (see also Wagner *et al*., 1999). Genetic or environmental stress mediated through the heat-shock protein Hsp90 allow the simultaneous release of cryptic genetic variation in a number of gene-transduction pathways. Intriguingly, they may have identified a molecular mechanism for linking the release of genetic variation to environmental change.

We thank Günter Wagner and members of his lab for discussions, and Bruce Walsh, Laurence Hurst and the reviewers for suggestions and comments on the manuscript. Chi-Hua Chiu is the recipient of an NSF/Sloan Foundation Postdoctoral Research Fellowship in Molecular Evolution, TFH was supported by the Norwegian Research Council, and AC was supported by a Department of Defense Graduate Research Fellowship.

REFERENCES

BARTON, N. & ROUHANI, S. (1987). The frequency of shifts between alternative equilibria. *J. theor. Biol.* **125,** 397–418.

BUNN, H. F. & FORGET, B. G. (1986). *Hemoglobin: Molecular, Genetic, and Clinical Aspects.* Philadelphia: W. B. Saunders.

BÜRGER, R. & EWENS, W. (1995). Fixation probabilities of additive alleles in diploid populations. *J. Math. Biol.* **33,** 557–575.

CHIU, C.-H., SCHNEIDER, H., SCHNEIDER, M. P. C., SAMPAIO, I., MEIRELES, C., SLIGHTOM, J., GUMUCIO, D. L. & GOODMAN, M. (1996). Reduction of two functional γ-globin genes to one: and evolutionary trend in New World monkeys (infraorder Platyrrhini). *Proc. Natl Acad. Sci. U.S.A.* **93,** 6510–6515.

CHIU, C.-H., SCHNEIDER, H., SLIGHTOM, J. L., GUMUCIO, D. L. & GOODMAN, M. (1997). Dynamics of regulatory evolution in primate γ-globin gene clusters: *cis*-mediated acquisition of simian γ fetal expression patterns. *Gene* **205,** 47–57.

CHIU, C.-H., GREGOIRE, L., GUMUCIO, D. L., MUNIZ, J. A. P. C., LANCASTER, W. D. & GOODMAN, M. (1999). Model for the fetal recruitment of simian gamma-globin genes based on finding from two new world monkeys *Cebus apella* and *Callithrix jacchus* (platyrrhini, Primates). *J. Exper. Zool.* (*Mol. Dev. Evol.*) **285,** 27–40.

CLARK, A. G. (1994). Invasion and maintenance of a gene duplication. *Proc. Natl Acad. Sci. U.S.A.* **91,** 2950–2954.

COYNE, J. A., BARTON, N. H. & TURELLI, M. (1997). A critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51,** 643–671.

DAWKINS, R. (1996). *Climbing Mount Improbable*. London: Viking.

DOVER, G. A. (1992). Observing development through evolutionary eyes: a practical approach. *BioEssays* **14,** 281–287.

DOVER, G. A. (1993). Evolution of genetic redundancy for advanced players. *Curr. Opin. Genet. Dev.* **3,** 902–910.

ELDER, J. F. J. & TURNER, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart. Rev. Biol.* **70,** 297–320.

FITCH, D. H. A., BAILEY, W. J., TAGLE, D. A., GOODMAN, M., SIEU, L. & SLIGHTOM, J. L. (1991). Duplication of the γ-globin locus mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc. Natl Acad. Sci. U.S.A.* **88,** 7396–7400.

FORCE, A., LYNCH, M., PICKETT, F. B., AMORES, A., YAN, Y. L. & POSTLETHWAIT, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545.

GELBART, W. M. & CHOVNIK, A. (1979). Spontaneous unequal exchange in the rosy region of *Drosophila melanogaster*. *Genetics* **92,** 849–859.

GERHART, J. & KIRSCHNER, M. (1997). *Cells, Embryos and Evolution: towards a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability*. Oxford: Blackwell.

GILLESPIE, J. H. (1984). Molecular evolution over the mutational landscape. *Evolution* **38,** 1116–1129.

GORDON, R. (1994). Evolution escapes rugged fitness landscapes by gene or genome doubling: the blessing of higher dimensionality. *Comput. Chem.* **18,** 325–331.

GOULD, S. J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. New York: Norton.

HOFFMANN, A. A. & PARSONS, P. A. (1997). *Extreme Environmental Change and Evolution*. Cambridge: Cambridge University Press.

HUGHES, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. B* **256,** 119–124.

HURST, L. D. & SMITH, N. G. C. (1998). The evolution of concerted evolution. *Proc. R. Soc. Land. B* **265,** 121–127.

INOMATA, N. & YAMAZAKI, T. (1996). Adaptive evolution at the molecular level of the duplicated *Amy* system in Drosophila. *J. Genet.* **75,** 125–137.

JOHNSON, R. M., BUCK, S., CHIU, C.-H., SCHNEIDER, H., SAMPAIO, I., GAGE, D. A., SHEN, T., SCHNEIDER, M. P. C., MUNIZ, J. A., GUMUCIO, D. L. & GOODMAN, M. (1996). Fetal globin expression in New World monkeys. *J. Biol. Chem.* **271,** 14 684–14 691.

KAUFFMAN, S. A. (1993). *The Origins of Order. Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

LANDE, R. (1985). Expected time for random genetic drift of a population between stable phenotypic states. *Proc. Natl Acad. Sci. U.S.A.* **82,** 7641–7645.

LANDE, R. & KIRKPATRICK, M. (1988). Ecological speciation by sexual selection. *J. theor. Biol.* **133,** 85–98.

LEWONTIN, R. C. (1978). Adaptation. *Sci. Am.* **239,** 212–231.

LISKAY, R. M. & STACHELEK, J. L. (1983). Evidence for intrachromosomal gene conversion in cultured mouse cells. *Cell* **35,** 157–165.

LISKAY, R. M., LETSOU, A. & STANCHELEK, J. L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115,** 161–167.

MARTIN, D. I. K., FIERING, S. & GROUDINE, M. (1996). Regulation of γ-globin gene expression: straightening out the locus. *Curr. Opin. Genet. Dev.* **6,** 488–495.

MORENO, G. (1994). Genetic architecture, genetic behavior, and character evolution. *Ann. Rev. Ecol. Syst.* **25,** 31–45.

NOWAK, M. A., BOERLIJST, M. C., COOKE, J. & MAYNARD SMITH, J. (1997). Evolution of genetic redundancy. *Nature* **388,** 167–171.

OHNO, S. (1970). *Evolution by Gene Duplication*. Heidelberg: Springer-Verlag.

OHTA, T. (1987). Simulating evolution by gene duplication. *Genetics* **115,** 207–213.

OHTA, T. (1989). The mutational load of a multigene family with uniform members. *Genet. Res.* **53,** 141–145.

OHTA, T. (1998). On the pattern of polymorphisms at major histocompatibility complex loci. *J. Mol. Evol.* **46,** 633–638.

OHTA, T. & DOVER, G. A. (1984). The cohesive population genetics of molecular drive. *Genetics* **108,** 501–521.

PHILLIPS, P. C. (1996). Waiting for compensatory mutation: phase zero of the shifting-balance process. *Genet. Res.* **67,** 271–285.

PRICE, T., TURELLI, M. & SLATKIN, M. (1993). Peak shifts produced by correlated response to selection. *Evolution* **47,** 280–290.

RAFF, R. A. (1996). *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. Chicago: University of Chicago Press.

RUBINITZ, J. & SUBRAMI, S. (1986). Extrachromosomal and chromosomal gene conversion in mammalian cells. *Mol. Cell. Biol.* **4,** 2253–2258.

RUTHERFORD, S. & LINDQUIST, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396,** 336–342.

SHAPIRA, S. K. & FINNERTY, V. G. (1986). The use of genetic complementation in the study of eukaryotic macromolecular evolution: rate of spontaneous gene duplication at two loci of *Drosophila melanogaster*. *J. Mol. Evol.* **23,** 159–167.

SLATKIN, M. (1986). Interchromosomal biased gene conversion, mutation and selection in a multigene family. *Genetics* **112,** 681–698.

SPOFFORD, J. B. (1969). Heterosis and the evolution of duplications. *Am. Nat.* **103,** 407–432.

ST ONGE, L., BOUCHARD, L. & BASTIN, M. (1993). High-frequency recombination mediated by polyomavirus large T antigen defective in replication. *J. Virol.* **67,** 1788–1795.

SÆTRE, G.-P. (2000). Sexual signals speciation. In: *Animal Signals: Signaling and Signal Design in Animal Communication* (Espmark, Y., Amundsen, T. & Rosenquist, G., eds), Trondheim: Tapir Academic Press.

WAGNER, A., WAGNER, G. P. & SIMILION, P. (1994). Epistasis can facilitate the evolution of reproductive isolation by peak shifts: a two-locus two-allele model. *Genetics* **138,** 533–545.

WAGNER, G. P., BOOTH, G. & BAGHERI-CHAICHIAN, H. (1997). A population genetic theory of canalization. *Evolution* **51,** 329–347.
WAGNER, G. P., CHIU, C.-H. & HANSEN, T. F. (1999). Is Hsp90 a regulator of evolvability? *J. Exp. Zool.* (*Mol. Dev. Evol.*) **285,** 116–118.
WALSH, J. B. (1987). Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* **117,** 543–557.
WALSH, J. B. (1995). How often do duplicated genes evolve new functions? *Genetics* **139,** 421–428.
WHITLOCK, M. C. (1995). Variance-induced peak shifts. *Evolution* **49,** 252–259.
WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16,** 97–159.
WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Cong. Genet*, Ithaca. Vol. 1, pp. 356–366.

Appendix A

The Rate and Probability of Peak Shift

Consider an ensemble of genes or populations that are presented with the problem of crossing the two-step adaptive valley illustrated in Fig. 1. The members of this ensemble can then be in any of the states indicated in Fig. 2. Let x_i be the probability that any one member of the ensemble is in state i, where i can be 1, 11, 12, 13, 14, or 44. We can then represent the model by the following set of differential equations:

$$\frac{dx_1}{dt} = -\delta x_1 + u_0(x_{11} + x_{12} + x_{13} + x_{14}), \quad (A.1)$$

$$\frac{dx_{11}}{dt} = \delta x_1 - (v_{12} + u_0)x_{11} + \kappa(x_{12} + x_{13} + x_{14}), \quad (A.2)$$

$$\frac{dx_{12}}{dt} = v_{12}x_{11} - (v_{23} + \kappa + u_0)x_{12}, \quad (A.3)$$

$$\frac{dx_{13}}{dt} = v_{23}x_{12} - (v_{34} + \kappa + u_0)x_{13}, \quad (A.4)$$

$$\frac{dx_{14}}{dt} = v_{34}x_{13} - (4sN\kappa + \kappa + u_0)x_{14}, \quad (A.5)$$

$$\frac{dx_{44}}{dt} = 4sN\kappa x_{14}. \quad (A.6)$$

The rationale for these equations is given in the main text. Provided the x_i's sum to 1 in the initial condition, their sum remains equal to 1. Eventually, all the probability mass accumulates in x_{44}. We are interested in computing the equilibrium rate, or flux, at which x_1 is converted to x_{44}. Call this rate λ, and assume that it is constant in time. To obtain this rate we add an arbitrary rate of conversion, β, from x_{44} to x_1 [i.e. add $-\beta x_{44}$ to eqn (A.6) and $+\beta x_{44}$ to eqn (A.1)]. In this way, the system has an equilibrium solution with $x_1 > 0$. At equilibrium it must be true that

$$\lambda = \beta \frac{x_{44}}{x_1} \quad (A.7)$$

and we proceed to compute x_{44}/x_1. Observe first that eqns (A.3)–(A.5), and (A.6) with $-\beta x_{44}$ added can be used to form the ratios x_{44}/x_{14}, x_{14}/x_{13}, x_{13}/x_{12} and x_{12}/x_{11}. Multiplying these together yields x_{44}/x_{11}, and using them together with eqn (A.2) computes x_{11}/x_1. Multiplying x_{44}/x_{11} and x_{11}/x_1 yields x_{44}/x_1. Using this in eqn (A.7) yields

$$\lambda = \frac{4sN\kappa v_{12}v_{23}v_{34}\delta}{u_0(u_0 + \kappa + 4sN\kappa)((u_0 + \kappa)^2 + (u_0 + \kappa)(v_{12} + v_{23} + v_{34}) + v_{12}v_{23} + v_{12}v_{34} + v_{23}v_{34}) + v_{12}v_{23}v_{34}(u_0 + 4sN\kappa)}, \quad (A.8)$$

which turns out to be independent of β. Thus, provided that there is a steady flux from state 1 to state 44, this flux is given by eqn (A.8).

This rate can be used to compute the probability, α, of peak shift given that a duplication has occurred, as it is true that $\alpha = \lambda/\delta$. An independent computation of α based on the Laplace transform of the system is sketched in Appendix B.

Observe that eqn (A.8) is invariant to the position of the three v-parameters in the sense that it

remains unchanged if any two of them are swapped for each other. Thus, there is little loss of generality in assuming that the v-parameters are identical.

Appendix B

The Mean Time to Peak Shift

Let the random variable t be the time of arrival at the new peak (i.e. at stage 44) starting from the pre-duplication state. The cumulative probability distribution of t is $x_{44}(t)$. Let $f(t)$ be the probability density of t, and

$$G(z) = \int_0^\infty e^{-zt} f(t)\,dt \tag{B.1}$$

be the Laplace transform of this probability density. Let $F_{44}(z)$ be the Laplace transform of $x_{44}(t)$. Now, as $x_{44}(t) = \int_0^t f(\tau)\,d\tau$, we have the relationship $G(z) = zF_{44}(z)$. We proceed to compute $F_{44}(z)$ by solving the Laplace transform of the linear system (A.1–A.6). If we write the linear system in vector notation as $dx(t)/dt = Ax(t)$, with initial condition $x(0) = \{x_1(0), x_{11}(0), x_{12}(0), x_{13}(0), x_{14}(0), x_{44}(0)\} = \{1, 0, 0, 0, 0, 0\}$, the vector of Laplace transforms of $x(t)$ is $F(z) = -(A - zI)^{-1}x(0)$, where I is the identity matrix. This can be computed by use of Mathematica and $F_{44}(z)$ is the last element of $F(z)$. Through differentiation of $G(z)$ we can now compute moments of the distribution of arrival times. The mean time to arrive at the new peak is

$$\bar{t} = -\lim_{z\to 0} G'(z)$$

$$= \frac{(u_0+\delta)(u_0+(1+4sN)\kappa)((u_0+\kappa)^2+3v(u_0+\kappa+v))+v^3(u_0+\delta+4sN\kappa)}{4sNv^3\kappa\delta}. \tag{B.2}$$

The method of Laplace transforms can also be used as an independent check of the computation of the fixation probability, α, given in Appendix A. We now work with the system where a duplication is assumed fixed in the population, i.e. we delete eqn (A.1) from the system, and have $x_{11}(0) = 1$ and $x_{12}(0) = x_{13}(0) = x_{14}(0) = x_{44}(0) = 0$ as initial conditions. When eqn (A.1) is deleted, the system "leaks" and the remaining probability mass $\alpha = \lim_{t\to\infty} x_{44}(t)$ is the probability that any given duplicate ends up at the new peak. It can be shown that $\lim_{t\to\infty} x_{44}(t) = \lim_{z\to 0} G(z)$, and computing the limit on the right-hand side with Mathematica confirms our result in Appendix A.

Appendix C

Peak Shifts by Genetic Drift

The genetic-drift model explained in the main text is given by the set of differential equations

$$\frac{dx_1}{dt} = -a_{12}x_1 + a_{21}x_2, \tag{C.1}$$

$$\frac{dx_2}{dt} = a_{12}x_1 - (a_{21}+a_{23})x_2 + a_{32}x_3, \tag{C.2}$$

$$\frac{dx_3}{dt} = a_{23}x_2 - (a_{34}+a_{32})x_3, \tag{C.3}$$

$$\frac{dx_4}{dt} = a_{34}x_3, \tag{C.4}$$

where x_i is the probability of being in state i, and a_{ij} represent a rate of transfer from state i to state j. By using the same technique as explained in Appendix A, it can be shown that the rate of transfer from state 1 to state 4 is

$$\lambda_d = \frac{a_{12}a_{23}a_{34}}{a_{21}a_{32}+a_{21}a_{34}+a_{23}a_{34}}. \tag{C.5}$$

Now, with the assumptions in the main text $a_{12} = a_{23} = 2NvP_d(s_d, N)$, where $P_d(s_d, N)$ is the probability of fixing an additive deleterious mutation with selection coefficient s_d against the heterozygote as given in the main text. Furthermore, $a_{21} = a_{32} = 4s_dNv$ and $a_{34} = 4sNv$. Using this in eqn (C.5) gives eqn (6).

The expected time to cross the valley by genetic drift is computed by the same means as in Appendix B:

$$\bar{t}_d = \frac{a_{12}(a_{23}+a_{32}+a_{34})+a_{21}(a_{32}+a_{34})+a_{23}a_{34}}{a_{12}a_{23}a_{34}}, \tag{C.6}$$

which under the assumptions in the main text leads to eqn (7).

Appendix D

Rate of Occurrence of Double Mutants in Large Populations

In this appendix, we derive the stochastic equilibrium distribution of double deleterious mutants in mutation–selection balance. We assume random mating, and that N is large and constant. Let $X_1(t)$ and $X_2(t)$ be random variables representing the number of single and double deleterious alleles in the population at time t. Then the cumulant generating function (c.g.f.) of $X_1(t+1)$ is

$$C_1(z;t+1) = \ln(E[\exp(zX_1(t+1)]) = \ln\left(E\left[\exp\left(z\left(\sum_{i=1}^{X_1(t)} y_i + A_1\right)\right)\right]\right), \tag{D.1}$$

where y_i is the number of offspring from the i-th individual and A_1 is the number of new mutants. Let $Y_1(z)$ be the c.g.f. of y_i and assume that this is independent of $X_1(t)$. We further assume that A_1 is independent of X_1, and binomially distributed with parameters v_1 for the mutation rate per allele and $2N$ for the number of alleles that can mutate ($2N$ is an approximation to $2N - X_1$). Then $M_1(z) = 2N\ln(1-v_1+v_1e^z)$ is the c.g.f. of A_1. From eqn (D.1) we can compute $C_1(z;t+1) = C_1(Y_1(z);t) + M_1(z)$. Let $C_1(z)$ be the c.g.f. of X_1 at stochastic equilibrium. This is then given implicitly as

$$C_1(z) = C_1(Y_1(z)) + M_1(z). \tag{D.2}$$

Differentiating $C_1(z)$ with respect to z and solving for $z=0$ gives $C_1'(0) = M_1'(0)/(1-Y'(0))$. As $Y'(0)$ is the mean number of offspring from a deleterious allele, the selection coefficient against that allele is $s_{d1} = 1 - Y'(0)$. Thus, we get $E[X_1] = C_1'(0) = 2Nv_1/s_{d1}$.

Let $C_2(z)$ be the equilibrium c.g.f. of X_2. This is of the same form as $C_1(z)$ except that $2N$ is replaced by the random variable X_1 since the new mutations must occur on alleles that already carry the first mutation. Thus, $M_2(z) = \ln(E[E[\exp(zA_2)|X_1]] = \ln(E(\exp(X_1\ln(1-v_2+v_2e^z)) = C_1(\ln(1-v_2+v_2e^z))$, where v_2 is the per locus mutation rate for the second mutation. From this we get

$$C_2(z) = C_2(Y_2(z)) + C_1(\ln(1-v_2+v_2e^z)). \tag{D.3}$$

Differentiating $C_1(z)$ and $C_2(z)$ at $z=0$ and solving, we get $E[X_2] = C_2'(0) = C_1'(0)v_2/(1-Y_2'(0))$. Thus,

$$E[X_2] = \frac{2Nv_1v_2}{s_{d1}s_{d2}}, \tag{D.4}$$

where $s_{d2} = 1 - Y_2'(0)$ is the selection coefficient against the double mutant.

For a peak shift to occur, the advantageous mutation must appear on one of the double-mutant alleles and then go to fixation. This happens at a rate $2sX_2v_3$, where v_3 is the probability of occurrence per allele for this last mutation. In eqn (9) we give the expectation of this rate. For simplicity, we assume that all mutation rates are equal, i.e. $v = v_1 = v_2 = v_3$.

Appendix E

Simulations

Individual-based simulations were performed in the following manner. The initial population was completely homozygous for the "1" haplotype and the simulation ended when the "44" or "40" haplotype became fixed in the population. During each generation several mutation steps, several gene conversion steps, a duplication step and a combined selection/mating step were performed. The mutation, conversion and duplication steps involved summing the number of

chromosomes qualified for each event and multiplying this by a per chromosome probability of occurrence. A uniform pseudorandom number between 0 and 1 was then compared to the product, if less the event occurred in a single, randomly chosen, individual. This was repeated for every possible event. Each event could occur at most once per generation. After the first series of steps a combined selection/mating step occurred, two individuals were chosen at random (with replacement and weighted by relative fitnesses) and mated to produce an offspring. This process continued until the appropriate number of progeny was generated. Fixation of the ending genotype was tested for and if found not to have occurred the process was repeated. Upon completion, the number of generations taken was recorded and the entire simulation started anew.