Gene Conversion Facilitates Adaptive Evolution on Rugged Fitness Landscapes

Philip Bittihn*,†,1 and Lev S. Tsimring*,†,1
*BioCircuits Institute, University of California San Diego (UCSD), La Jolla, California 92093-0328 and †San Diego Center for Systems Biology, La Jolla, California 92093-0688
ORCID IDs: 0000-0002-1276-9381 (P.B.); 0000-0003-0709-3548 (L.S.T.)

ABSTRACT Gene conversion is a ubiquitous phenomenon that leads to the exchange of genetic information between homologous DNA regions and maintains coevolving multi-gene families in most prokaryotic and eukaryotic organisms. In this paper, we study its implications for the evolution of a single functional gene with a silenced duplicate, using two different models of evolution on rugged fitness landscapes. Our analytical and numerical results show that, by helping to circumvent valleys of low fitness, gene conversion with a passive duplicate gene can cause a significant speedup of adaptation, which depends nontrivially on the frequency of gene conversion and the structure of the landscape. We find that stochastic effects due to finite population sizes further increase the likelihood of exploiting this evolutionary pathway. A universal feature appearing in both deterministic and stochastic analysis of our models is the existence of an optimal gene conversion rate, which maximizes the speed of adaptation. Our results reveal the potential for duplicate genes to act as a "scratch paper" that frees evolution from being limited to strictly beneficial mutations in strongly selective environments.

KEYWORDS gene conversion; gene duplication; pseudogenes; fitness landscape; adaptive evolution

MUTAGENESIS and selection are the driving force of evolution. Among many types of mutations, gene duplication plays an important role in expanding the genotype and introducing novel gene functions in all life forms (Ohno 2013). Gene duplication first leads to the creation of an identical copy in the genome. The fate of the newly minted copy then varies: it can mutate and acquire new or improved functionality (neofunctionalization), two or more copies may genomically diverge and specialize in two subfunctions (subfunctionalization), or one of the copies can become nonfunctional (nonfunctionalization) (Lin *et al.* 2006; Plata and Vitkup 2014). After the duplication of a single gene, independent mutations can occur in each of the copies. However, several studies have found evidence for "concerted evolution" among members of the same gene family (Elder and Turner 1995), which results in stronger similarity between genes than expected from independent mutations. Concerted evolution is found in most organisms from bacteria to mammals, and can only be explained by a continuing exchange of genetic code between duplicate genes (Liao 1999), a process known as *gene conversion* (Chen *et al.* 2007). The most widely accepted mechanism of gene conversion is based on double-strand-break repair and invasion of a homologous sequence from a different locus (Hastings 2010). The molecular kinetic parameters of this process have only recently been quantified in a controlled setting, explicitly showing in a laboratory evolution experiment that gene conversion results in the replacement of a contiguous DNA section by a homologous sequence (Paulsson *et al.* 2017). On an abstract level, gene duplication and conversion proceed as illustrated in Figure 1A, with gene conversion counteracting the divergence of the sequences (Fawcett and Innan 2011).

It is widely believed that nonfunctionalization is by far the most common fate of gene duplicates, with 90% of newly duplicate genes quickly turning into "pseudogenes" (Lynch and Conery 2000; Plata and Vitkup 2014). However, rendering a gene silent does not stop the evolution of its DNA sequence. Moreover, gene conversion, because it is linked to DNA replication, does not depend on the functionality of the gene product (Paulsson *et al.* 2017). Therefore, neutral mutations in a nonfunctionalized pseudogene can be transferred into the functional gene. If gene conversion introduces

doi: https://doi.org/10.1534/genetics.117.300350
Manuscript received April 19, 2017; accepted for publication September 30, 2017; published Early Online October 4, 2017.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300350/-/DC1.
[1]Corresponding authors: Bio Circuits Institute, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0328. E-mail: pbittihn@ucsd.edu; and ltsimring@ucsd.edu

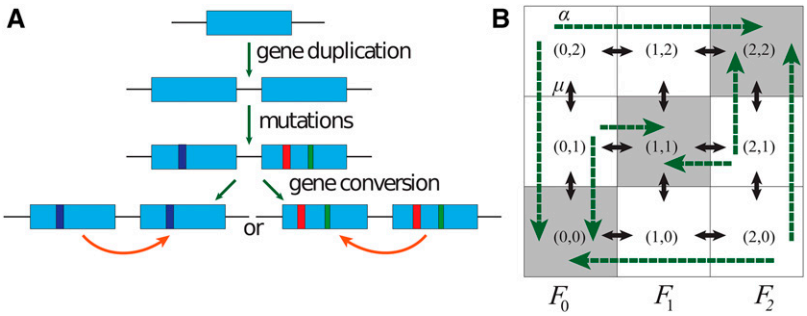

Figure 1 (A) Cartoon of the gene duplication and conversion. (B) Schematic diagram of the minimal 3 × 3 model. Short bidirectional arrows denote mutations and green arrows denote gene conversions.

a lethal or deleterious mutation into a functional gene, this mutation would be quickly purged from the population; however, a beneficial mutation would have a high probability of fixation. Even a nonfunctional gene can therefore potentially be the source of mutations for adaptation.

This effect may be particularly important if, in order to achieve a higher-fitness state, a gene has to accumulate a number of mutations that individually are deleterious. This evolutionary process is commonly described as an adaptive walk over a complex fitness landscape (Flyvbjerg and Lautrup 1992). When the population is at a local fitness peak, to acquire a new functionality or a higher fitness, it would need to traverse a valley of lower fitness through a series of correlated mutations. For example, Wu *et al.* (2016) recently showed that many direct paths in protein fitness landscapes are blocked by deleterious effects of intermediate steps, which may considerably slow down adaptive evolution. The theoretical description of population dynamics in nonmonotonous fitness landscapes ("fitness valley crossing") have been worked out in a number of publications over the last 20 years (Innan and Stephan 2001; Iwasa *et al.* 2004; Weinreich and Chao 2005; Fisher 2007; Serra and Haccou 2007; Weissman *et al.* 2009; Saakian *et al.* 2017). These studies show that, in the biologically relevant range of parameters, the "valley crossing" becomes extremely slow for highly deleterious intermediate states because of efficient purifying selection of the intermediate mutants. There is, however, an intriguing unexplored possibility that gene conversion between homologous sequences could provide a mechanism of genetic variation that is shielded from strong purifying selection if the original gene can fulfill its intended biological function while its copy explores the fitness landscape with little selective pressure. There is evidence suggesting that, indeed, duplicate or pseudogenes provide an important source of genetic variation for adaptive mutations (Hayakawa *et al.* 2005).

Several population genetics models of gene duplication and conversion have been introduced over the years (Walsh 2003; Kondrashov and Koonin 2004; Innan 2009; Innan and Kondrashov 2010). While most of these studies assumed that the original and duplicate copies of a gene evolve independently, some incorporated concerted evolution by including recurring gene conversion events that lead to "reverse mutations" in the paralogous genes, and, therefore, slow down their divergence (Teshima and Innan 2004). Furthermore, typical population genetic models assume that a gene's fate is determined by a single mutation, either rendering it nonfunctional, or, in rare cases, endowing it with a different or improved functionality. These models were mostly concerned with the fixation and function of newly duplicated genes, but some recent studies addressed the implications of gene conversion for the adaptive evolution of the original gene's function due to the increase of genetic diversity within multi-gene families (Takuno *et al.* 2008) or dosage effects (Mano and Innan 2008; Innan and Kondrashov 2010). However, there are currently no models that address the advantage that gene conversion can provide to adaptive evolution in complex fitness landscapes.

Here, we study theoretically the role of gene conversion in adaptive evolution on nontrivial fitness landscapes for the simplest case of one pair of duplicate genes. In the first part of the paper, using a stochastic population dynamics model in a minimal nontrivial fitness landscape, we explore how gene conversion accelerates the fixation of a beneficial mutation that is only accessible via a deleterious intermediate mutation. In the second part of the paper, extending the concept to more complex genotypes and landscapes inspired by the classical NK-model for epistasis (Kauffman 1993), we then analyze how the rate of adaptation depends on the rates of gene conversion and mutation and the population size, enabling us to determine the conditions for the optimal gene conversion rate in rugged fitness landscapes.

Minimal model

Consider a pool of individuals with a genome consisting of a single "gene" that can only have three states: 0, 1, and 2. We assume a constant population size N, so the individuals reproduce with the replication rate equal to their corresponding fitness values, and, to keep the population size constant, each replication event is accompanied by removal of a random individual from the population (Moran process). The population is initially monoclonal with all individuals in state 0. We assume that a single mutation can only switch the gene between states 0 and 1, or between states 1 and 2, but not between 0 and 2 directly. This would be the case, for example, if we consider a 2-bit string as a model gene, and a mutation can only flip one bit at a time (both one-bit mutants can be lumped in state 1). Evidently, only two sequential mutations may bring the system from the original state to the state with both bits flipped. We refer to the fitness values

of states 0, 1, and 2 as F_0, F_1, and F_2, respectively, and assume that they satisfy the condition $F_1 < F_0 < F_2$. Thus, to mutate from state 0 to a higher-fitness state 2, an individual has to cross a lower fitness "valley" of state 1. To describe gene conversion with a duplicate gene, we augment the "active" gene by a "passive" gene that can also be in any of the three states $\{0, 1, 2\}$, but the state of the passive copy does not affect the fitness of the individual. We denote the state of an organism in this 3×3 state space by (i,j) where the two indices $i,j \in \{0, 1, 2\}$ correspond to the types of the active and passive genes, respectively (Figure 1B), meaning that the fitness of the state (i,j) is F_i. We assume that mutations from state (i,j) occur only during replication, and so they happen with propensity μF_i (proportional to the fitness of the active copy) to each of the neighboring states, corresponding to forward and backward mutations in both active and passive copies. Gene conversion is modeled as copying the active gene into the passive gene $(i,j) \to (i,i)$, or back $(i,j) \to (j,j)$, with the rate αF_i. Without loss of generality, we can rescale time to fix the wild-type fitness at 1 and take $F_0 = 1, F_1 = 1 - \delta, F_2 = 1 + s$. In the following, we will always assume $\mu, \alpha \ll s \ll 1$.

We performed stochastic simulations of this model with fixed selective advantage $s = 0.1$ and different values of $\delta, \mu, \alpha,$ and N using the Gillespie algorithm (Gillespie 1977). We start simulations from all N individuals localized in state $(0,0)$, and observe the evolution of the distribution of the occupation numbers $X_{i,j}$ of each state (i,j) over time. For mutation rate $\mu = 10^{-3}$ and population size $N = 800$, mutations occur relatively frequently. Without gene conversion ($\alpha = 0$,) the state of the passive gene is irrelevant, and evolution proceeds via regular mutations along the state of the active gene (Figure 2A). The population spreads from state 0 to state 1, and then to 2, where it eventually localizes. The efficiency of this process depends critically on the fitness of the intermediate state 1, and for large δ (small F_1) is extremely low. However, for finite gene conversion rate ($\alpha > 0$), the population can reach state 2 for the active gene even across a deep fitness valley ($\delta = 1$), as shown in Figure 2B. The population spreads transiently to state (0,2) (light blue) by mutations $(0,0) \to (0,1) \to (0,2)$ and then crosses the fitness valley directly to state (2,2) (bright green) by gene conversion.

If the flux of mutations is sufficiently small, the dynamics look markedly different (Figure 2, C and D), as a polymorphic population similar to Figure 2, A and B cannot be maintained. Instead, each mutation either gets fixed or is lost, *i.e.*, the population homogenizes before the next mutation occurs, leading to "switching" dynamics. If a weakly deleterious mutation into state 1 gets fixed due to random genetic drift (red color), the population can reach state 2 even when $\alpha = 0$ (Figure 2C). For $\alpha > 0$ and $\delta = 1$ (Figure 2D), the population reaches state (2,2) (bright green) via conversion after a neutral mutation to state (0,2) becomes fixed in the population (light blue color).

As shown by Weissman *et al.* (2009), the way a population crosses intermediate states to a final beneficial mutant (in our case either along the direct, or along the gene conversion, path) can differ significantly depending on population size N and the depth of the fitness valley δ. It can range from (1) a sequence of fixation events at very small population sizes, with the population occupying essentially a single state at a time (Figure 2, C and D), to (2) "stochastic tunneling" at intermediate population sizes, when intermediate states can be occupied long enough to result in the production of a beneficial mutant (Figure 2, A and B), to (3) completely deterministic continuous generation of mutants at very large population sizes. A number of additional mixed cases can be expected when either the two paths (direct or via gene conversion) are in different regimes, or individual transitions within the paths have to be treated with different approximations. While all of these mixed regimes can in principle be analyzed, we will concentrate on the three pure regimes mentioned above. The exact population size requirements for each case will be specified below. We will start with stochastic tunneling, as it provides a convenient starting point for specifying population size limits for all three regimes.

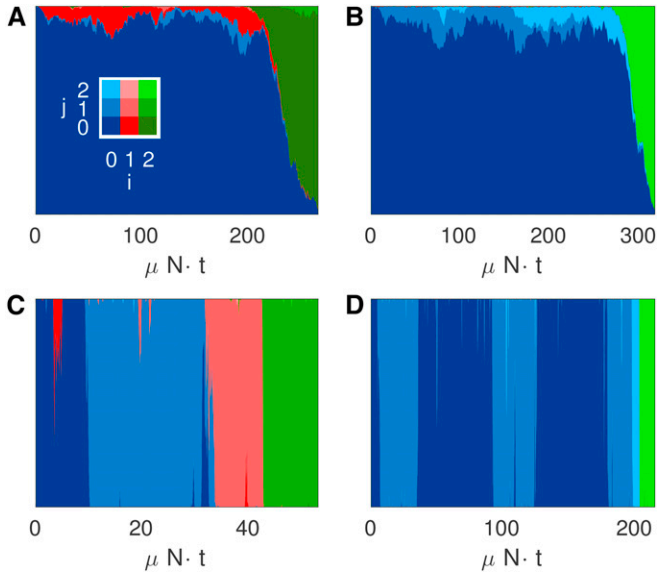

Figure 2 Stochastic simulations of the minimal gene conversion model. Stacked areas of different colors represent the number of individuals in each state according to the lookup table in (A). Colors blue, red, and green correspond to the state of the active gene determining the fitness, different shades of the same color indicate the state of the passive gene. Parameters: (A) $N = 800$, $\delta = 0.03$, $\mu = 10^{-3}$, $\alpha = 0$. (B) $N = 800$, $\delta = 1$, $\mu = 10^{-3}$, $\alpha = 10^{-3}$. (C) $N = 100$, $\delta = 0.01$, $\mu = 10^{-4}$, $\alpha = 0$. (D) $N = 100$, $\delta = 1$, $\mu = 10^{-4}$, $\alpha = 10^{-4}$. In all panels, $s = 0.1$.

Stochastic tunneling

In this regime, the mutant subpopulation remains much smaller than N until the final stage, when beneficial mutants are established in a population and finally grow to fixation (see Figure 2, A and B). The adaptive evolution proceeds via stochastic tunneling through the low-occupancy intermediate states. Weissman *et al.* (2009) presented the most comprehensive theoretical description of stochastic tunneling to a beneficial

state with K mutations through a low-fitness valley of $K-1$ deleterious or neutral mutations. Their approach is based on the notion of a "successful" individual. An individual is considered successful if one of its descendants will mutate into the final (beneficial) state and get fixed in the population. We define the fixation time of the beneficial mutation as the time when half of the population has reached the state 2. In the stochastic tunneling regime, the dominant contribution to this fixation time is the time until a successful individual is produced. This time is inversely proportional to the rate of producing a single mutant and to the probability p_1 that this mutant is successful. The latter can be computed by the first-step analysis described in Appendix C of Weissman *et al.* (2009).

We can apply this approach to our problem by computing the fixation times of the beneficial state 2 for evolution via two separate paths, which, in the first approximation (*i.e.*, for small mutation and conversion rates), can be considered independent. The direct path ($0\to1\to2$) is essentially the valley crossing problem of Weissman *et al.* (2009) for $K=2$ (with one deleterious intermediate mutation), with a few minor differences. First, our model postulates that the fitness of an individual characterizes its *total* division rate, which includes both normal symmetric binary divisions and mutations. Accordingly, the mutation rate of the individual is equal to the product of the nominal per-generation mutation rate μ times its fitness. Additionally, we take into account back mutations $1\to0$ and $2\to1$ with the same rates as the forward mutations. Still, the first-step analysis can be applied here, and in the limit of strongly deleterious intermediate state ($\delta \gg \sqrt{(1-\delta)\mu s}$), the rate of producing single mutants in state 1 is $N\mu$, and the probability that this mutant is successful is $p_1=(1-\delta)\mu s/\delta$, so the approximate expression for the fixation time $T_1=1/(\mu N p_1)$ of the direct path reads

$$T_1 \approx \frac{\delta}{N\mu^2 s(1-\delta)} \tag{1}$$

in agreement with previous studies of fitness valley crossing by Innan and Stephan (2001), Weissman *et al.* (2009) (see Supplemental Material, File S1 for more details). As expected, the fixation time is inversely proportional to the population size, and strongly dependent on the mutation rate μ and the selective advantage of the beneficial state s.

The second path that involves gene conversion can be exactly isolated by assuming $F_1=0$ ($\delta\to1$). In this case, the mutants in the intermediate state 1 do not reproduce. and are effectively excluded from the evolutionary process. Thus, the only path to the beneficial state 2 is $(0,0)\to(0,1)\to(0,2)\to(2,2)$, where the first two steps are neutral mutations in the passive copy, and the last step is the gene conversion from the passive to the active copy. Although this regime is similar to the case $K=3$, with two neutral intermediate mutations in Weissman *et al.* (2009), the inclusion of back mutations [$(0,1)\to(0,0)$ and $(0,2)\to(0,1)$] and—more importantly—back conversions [$(0,1)\to(0,0),(0,1)\to(1,1),(0,2)\to(0,0)$] leads to significantly different results: using the first-step method again, we can obtain the full solution for the probability p_{01} that the (0,1) mutant is successful (see File S1) and derive the fixation time T_2 as $(N\mu p_{01})^{-1}$. This fixation time is plotted in Figure 3A as a function of α for fixed N and μ. As seen from this figure, it is nonmonotonic, *i.e.*, decreases at small α and increases at large α, and there is an optimal $\alpha_m \approx 2^{-4/3}(s\mu^2)^{1/3}$ for which the fixation time T_2 is minimized. It is easy to understand the origin of this optimum: for small α the gene conversion proceeds very slowly, and for large α the process is hampered by the strong flux of mutants from states (0,1) and (0,2) back to state (0,0) and to the low-fitness state (1,1), thus reducing the flux into state (2,2). In other words: gene conversion enables the transfer of the beneficial genotype 2 from the passive to the active gene and is thus beneficial, but only up to a certain rate. Further increasing the gene conversion rate prevents the passive copy from reaching the beneficial state, because it is either replaced by the active copy, or a deleterious intermediate state of the passive copy is transferred to the active gene. This trade-off, which is due to the bidirectionality of gene conversion, is a universal effect encountered in all the regimes and scenarios considered in this study.

The approximate formulas for p_{01} for small and large α can be obtained in closed form (see File S1),

$$p_{01} \approx \begin{cases} \sqrt{\mu\sqrt{\alpha s}}, & \alpha \ll (s\mu^2)^{1/3} \\ \dfrac{\mu}{2}\sqrt{\dfrac{s}{\alpha}}, & \alpha \gg (s\mu^2)^{1/3} \end{cases}, \tag{2}$$

which yields the following expressions for the fixation time T_2 of the gene conversion path,

$$T_2 \approx \begin{cases} 1\Big/\left[N\mu^{3/2}(\alpha s)^{1/4}\right], & \alpha \ll (s\mu^2)^{1/3} \\ 2\alpha^{1/2}\big/\left[N\mu^2 s^{1/2}\right], & \alpha \gg (s\mu^2)^{1/3}. \end{cases} \tag{3}$$

The two cases for large and small α connect at $\alpha_m = 2^{-4/3}(s\mu^2)^{1/3}$; however, for specifying approximate regime boundaries, we omit the numerical factor $2^{-4/3}$. Note that the smaller-α regime corresponds to the approximation derived by Weissman *et al.* (2009), whereas the second, larger-α, regime is a result of back conversions and conversions to the deleterious state (1,1) that lead to a reduction of the overall success probability of mutants in state (0,1). Note also that there is a third regime for very small $\alpha \ll \mu^2/s$ (see File S1) which is usually irrelevant because p_{01} drops very quickly in this regime, and the direct path will consequently provide a shorter fixation time.

For small but nonzero fitness F_1 of the intermediate mutant, when both direct and gene conversion paths can lead to fixation of state 2 in comparable times, the overall fixation time can be found as $T=(T_1^{-1}+T_2^{-1})^{-1}$, since these paths are nonoverlapping for large δ and small α. An example for the overall fixation time of the beneficial mutation for fixed N and μ is shown in Figure 3B for a range of conversion rates. For large α, the approximation $T=(T_1^{-1}+T_2^{-1})^{-1}$ slightly overestimates the fixation time because of the neglected interaction between the two paths: conversions from (0,1) to

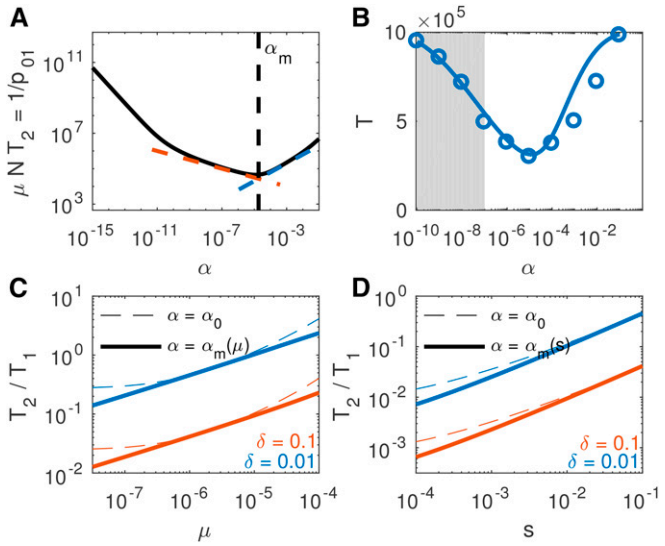

Figure 3 Stochastic tunneling regime of the minimal model. (A) Normalized fixation time T_2 along the gene conversion path, only. Colored dashed lines indicate the approximations of Equation (2) for small (red) and large α (blue). The vertical dashed line indicates the optimal gene conversion rate $\alpha_m \approx (s\mu^2)^{1/3}$. Parameters are $\mu = 10^{-6}$ and $s = 0.1$. (B) Numerical simulations (circles) and theory according to $T = (T_1^{-1} + T_2^{-1})^{-1}$ for the full minimal model for $N = 10^5$, $\mu = 10^{-6}$, $\delta = 0.01$, and $s = 0.1$. The shaded area indicates the gene conversion rates at which the tunneling approximation for the gene conversion path becomes formally invalid. (C) Fixation time reduction T_2/T_1 provided by gene conversion as a function of μ for $s = 0.1$ and two different values of δ (the ratio is independent of N), assuming the population size is in the stochastic tunneling regime. The solid lines were calculated with the optimal $\alpha_m = 2^{-4/3}(s\mu^2)^{1/3}$, the dashed lines use fixed $\alpha_0 = (16 \cdot 10^{13})^{-1/3}$ (the optimal α for the parameters in B). (D) Same as in (C), but for varying selective advantage s with fixed mutation rate $\mu = 10^{-6}$.

the deleterious state (1,1) can facilitate tunneling along the direct path to the better fitness state (2,1) instead of simply representing an effective reduction in fitness.

Substituting the optimal α_m into Equation (3) shows that the minimal fixation time achievable via gene conversion is $T_{2,\min} \approx 2^{1/3}(Ns^{1/3}\mu^{5/3})^{-1}$. Comparing this expression with the fixation time T_1 via the direct path, Equation (1), we observe that it is also inversely proportional to N, but depends more weakly on the mutation rate μ, and, much more weakly, on the selective advantage s. Therefore, as shown in Figure 3, C and D, the speedup through gene conversion in the stochastic tunneling regime is expected to be larger for smaller μ and s, as $T_{2,\min}$ grows more slowly than T_1, while it does not change with N. The scaling behavior is reversed and becomes more favorable for the direct path if the fitness valley is effectively neutral ($\delta \ll \sqrt{(1-\delta)\mu s}$, see File S1), which is intuitive since the conversion path involves an extra step.

The applicability of the stochastic tunneling calculation requires that mutant subpopulations remain smaller than N at all times, except for the final fixation state of the beneficial mutant. For the direct path, the sufficient condition for this is $p_1 \gg \rho_1$, where p_1 is the probability to produce a successful single mutant, and ρ_1 is the probability of fixation in state 1. For the strongly deleterious intermediate state ($\delta \gg 2\sqrt{(1-\delta)\mu s}$), we use Kimura's diffusion approximation to the fixation probability (Kimura 1962; Otto and Whitlock 2006) for a mutation with selective advantage s

$$p(s) = \frac{1 - \exp(-s)}{1 - \exp(-Ns)}. \tag{4}$$

It is exponentially small for negative s (deleterious mutations), equal to $1/N$ for $s = 0$, and is $\approx s$ for small $s > 0$ and large Ns. In this case, therefore, $\rho_1 = p(-\delta)$. This translates to the condition

$$N \gg \delta^{-1} \log\left(1 + \delta \frac{e^{\delta} - 1}{\mu(1-\delta)s}\right) \tag{5}$$

which corresponds to Equation (31) of Weissman *et al.* (2009). For the gene conversion path, the corresponding condition is $p_{01} \gg \rho_{01}$, where ρ_{01} is the fixation probability in state (0,1). For this state, gene conversion to the states (1,1) and (0,0) effectively reduces the division rate by a factor $1 - 2\alpha$, so its fixation probability is given by $\rho_{01} = p(-2\alpha)$. Substituting p_{01} from Equation (2), we arrive at

$$N \gg \begin{cases} \dfrac{1}{\sqrt{\mu\sqrt{\alpha s}}}, & \alpha \ll (s\mu^2)^{1/3} \\ \dfrac{1}{2\alpha} \log\left(1 + \dfrac{\alpha}{\mu}\sqrt{\dfrac{\alpha}{s}}\right), & \alpha \gg (s\mu^2)^{1/3}, \end{cases} \tag{6}$$

(see File S1). In Figure 3B, the shaded area indicates the values of α at which this condition is violated for the gene conversion path, in this case only $N < 1/\sqrt{\mu\sqrt{\alpha s}}$. However, the direct path dominates the dynamics in this area, and, so, despite this, the quantitative deviations of the theoretical predictions remain invisible.

Besides a minimum population size, another requirement for the validity of the stochastic tunneling approximation is that N is *not too large*, so the lineages of successful first mutants do not overlap, and the first one always leads to fixation. The sufficient condition for this to hold is

$$N \ll \delta^2 / [s\mu^2(1-\delta)] \tag{7}$$

for the direct path, with strongly deleterious intermediate state 1 and

$$N \ll \begin{cases} \dfrac{1}{\mu}, & \alpha \ll (s\mu^2)^{1/3}, \\ \dfrac{4\alpha^{3/2}}{\mu^2 s^{1/2}}, & \alpha \gg (s\mu^2)^{1/3}. \end{cases} \tag{8}$$

for the gene conversion path (see File S1 for details).

Continuous evolution in large populations

When the conditions discussed at the end of the previous section are violated, the stochastic tunneling approximation is

no longer valid. For very large populations (the validity conditions will be specified below), the population continuously spreads to neighboring states, and we can analytically describe the evolutionary dynamics by using deterministic mass-action equations for the occupation numbers $X_{i,j}$:

$$\begin{aligned}\dot{X}_{i,j} \;=\; & F_i X_{i,j} - d X_{i,j} + \mu F_{i-1} X_{i-1,j} + \mu F_i X_{i,j-1} + \mu F_{i+1} X_{i+1,j} \\ & + \mu F_i X_{i,j+1} - (4 - \delta_{i0} - \delta_{j0} - \delta_{i2} - \delta_{j2}) \mu F_i X_{i,j} \\ & - 2\alpha F_i X_{i,j} + \alpha \sum_{k=0}^{2} \left(F_i X_{i,k} + F_k X_{k,i}\right) \delta_{ij}, \{i,j\} = 0,1,2\end{aligned} \tag{9}$$

where δ_{ij} is the Kroneker delta, and $X_{i,j} = 0$ for either i or $j < 0$ or > 2. To mimic the Moran process deterministically, we set the death rate d equal to the current average fitness of the whole population, $d = \sum_{i,j=0}^{2} F_i X_{i,j}/N$, where $N = \sum_{i,j=0}^{2} X_{i,j}$ is the constant population size.

For model (9), the fixation time can be found analytically in two limiting cases using Laplace transforms (see File S1 for derivations). In the absence of gene conversion ($\alpha = 0$), and for small mutation rate $\mu \ll s$, the fixation time via the direct path is

$$T_1 \approx \frac{1}{s} \log\left[\frac{s(\delta + s)}{2\mu^2(1-\delta)}\right] \tag{10}$$

In the other limiting case, we again consider only the evolutionary path to state 2 via gene conversion by assuming $\alpha > 0$, but taking $F_1 = 0$. In this case, the fixation time is given by the formula

$$T_2 \approx \frac{1}{s} \log\left[\frac{s(s+2\alpha)^2}{2\alpha\mu^2}\right] \tag{11}$$

According to Equation (11), the switching time T_2 is again a nonmonotonous function of α with a minimum at $\alpha_m = s/2$. We can also find the value of α at which $T_1 = T_2$, as a root of the quadratic equation $(1-\delta)(s+2\alpha)^2 = \alpha(s+\delta)$. For small $\alpha \ll s$ the solution is

$$\alpha_c = \frac{(1-\delta)s^2}{s+\delta}. \tag{12}$$

For $\alpha \ll \alpha_c$, the effect of gene conversion is negligible, and the fixation time is given by T_1, which depends on F_1. For $\alpha \gg \alpha_c$, gene conversion dominates the evolution, and the fixation time is given by T_2, which is independent of δ. Interestingly, the crossover point α_c is independent of the mutation rate μ.

Simulations of the full deterministic model (9) are in agreement with these results. Figure 4 shows the fixation time of state 2 as a function of α for different values of δ and μ. As predicted by the theory, the fixation time is approximately T_2 for large enough α, but, at a certain α_c, it crosses over to the α-independent, but δ-dependent value T_1. The fixation times deviate somewhat from the analytical estimates, since we assume small $X_{i,j}$ for all $(i,j) \neq (0,0)$ to derive them (see File S1 for details), but use a numerical fixation criterion of 50% for state 2. However, the relative numerical error of our analytical predictions is small and the parameter dependence of the fixation times is accurately described by the theory.

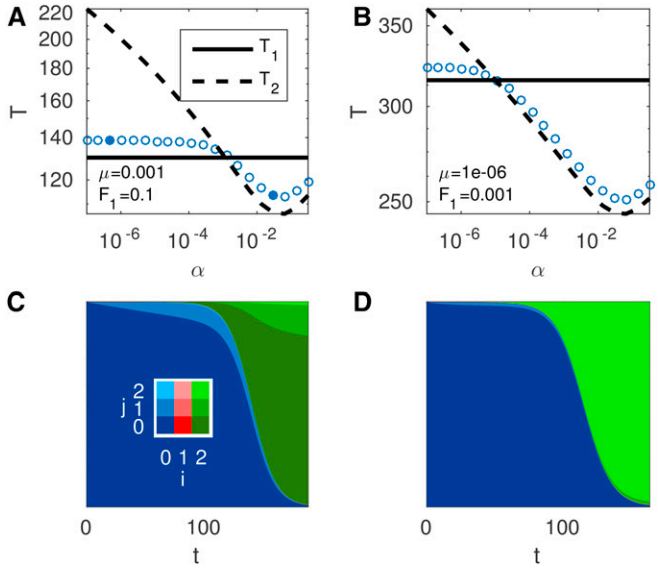

Figure 4 Theory and simulations for deterministic system [9] with $s = 0.1$. (A) Numerical simulations (symbols) compared to the theoretical predictions of T_1 and T_2 from Equations (10), (11) (lines) for $\mu = 10^{-3}$ and $\delta = 0.9$ and a range of conversion rates α. (B) Same as in (A), but for smaller $\mu = 10^{-6}$ and $\delta = 0.999$. (C) Trajectory of (9) for $\mu = 0.001$, $\delta = 0.9$, and $\alpha = 4.83 \times 10^{-7}$ (left filled circle in A). The system takes the direct path to reach higher fitness state 2 via state 1, even though the fraction of the population in state 1 is minuscule and the red shades are not visible. (D) Trajectory of (9) for $\mu = 10^{-3}$, $\delta = 0.9$, and $\alpha = 0.03$ (right filled circle in A). The system reaches state 2 through the conversion process $(0,2) \rightarrow (2,2)$.

Intuitively, a smaller mutation rate μ leads to longer fixation times across all α (Figure 4B), as can also be seen directly from Equations (10) and (11). For fixed s, the crossover value α_c decreases monotonically with increasing δ, and so the range of conversion rates α with shortened fixation times (when $T_2 < T_1$) increases (*cf.* Figure 4, A and B). For $\alpha < \alpha_c$, the high fitness state is reached directly through mutations, such that state (2,0) is populated first (Figure 4C). In contrast, for $\alpha > \alpha_c$, a state with beneficial fitness is reached by gene conversion into state (2,2) (Figure 4D), leading to reduced fixation time (Figure 4A).

The deterministic description presented in this section is applicable if for every transition between states included in the model $p_{a\to b} X_a(t_e) \gg 1$, where $p_{a\to b}$ is the probability of transition from state a to state b, and $X_a(t_e)$ is number of individuals of type a at the time t_e at which the population of the beneficial mutants is established, *i.e.*, reaches $O(s^{-1})$ (see Fisher 2007; Weissman *et al.* 2009). For the direct path with strongly deleterious state 1, the most stringent limitation is imposed by the transition from state 1 to 2, for which $p_{1\to 2} = \mu(1-\delta)$ and $X_1(t_e) = \mu N/\delta$, so the deterministic

description is valid for $N \gg \delta/[\mu^2(1-\delta)]$ [*cf.* Equation (36) of Weissman *et al.* (2009)]. For the gene conversion path, the most stringent limitation is imposed by the transition from state (0,2) to (2,2), for which $p_{02\to 22} = \alpha$ and $X_{02}(t_e) = N\mu^2 s^{-2}\log^2(s^2/N\alpha\mu^2)$ (see File S1). Then, the condition of applicability of the fully deterministic description of the gene conversion path is

$$N \gg \frac{s^2}{\alpha\mu^2}. \tag{13}$$

It is easy to see that there is a gap between this fully deterministic regime and the stochastic tunneling regime ($N \ll \mu^{-1}$). In this gap, various intermediate regimes apply, in which some transitions can be treated deterministically, while others still exhibit stochastic tunneling. The corresponding theoretical approximations can be developed for these scenarios, but since there are many cases to consider for different combinations of parameters, we will not describe them here.

Monoclonal regime in small populations

If N becomes so small that the conditions $p_1 \gg \rho_1$ (direct path) or $p_{01} \gg \rho_{01}$ (conversion path) for the applicability of the stochastic tunneling calculation are violated, the intermediate mutants may get fixed before giving rise to the beneficial mutant. For the direct path, the population size has to be extremely small for the strongly deleterious mutation to have a chance of getting fixed before a successful single mutant is produced by tunneling, because ρ_1 is exponentially small for $\delta N \gg 1$. However, if, nonetheless $p_1 \ll \rho_1$, *i.e.*,

$$N \gg \delta^{-1}\log\left(1 + \delta\frac{e^{\delta}-1}{\mu(1-\delta)s}\right) \tag{14}$$

[the reverse of Equation (5)] is fulfilled, the intermediate state (1,*) will always get fixed before the mutation to state (2,*) happens. We call this the *monoclonal regime*.

For the conversion path with two intermediate mutations, the condition for fully monoclonal regime is not simply the reverse of the stochastic tunneling condition $p_{01} \ll \rho_{01}$, since it only guarantees the fixation of state (0,1) but not of state (0,2). The condition for the latter is $p_{02} \ll \rho_{02}$, which is always more stringent. Substituting $p_{02} = \sqrt{\alpha s}$ and $\rho_{02} = p(-\alpha)$, we obtain

$$N \ll \frac{1}{\sqrt{\alpha s}} \tag{15}$$

(see File S1 for details). In the gap between this condition and (6), there is a semimonoclonal regime, where (0,1) mutants get fixed, and the population then stochastically tunnels through (0,2) to the final beneficial state (2,2). As mentioned in the Introduction, we ignore these mixed scenarios in this study.

If we assume that the population gets fixed at every intermediate state along the way before proceeding (or possibly reverting due to back mutations or conversions), we can convert the full minimal model to a continuous-time Markov chain. In this *monoclonal approximation*, different states represent homogeneous populations with a certain genotype and the transition probabilities between states depend on the corresponding mutation rates and fixation probabilities. This is similar to the "sequential fixation" regime in Weissman *et al.* (2009), but in our case the states are not necessarily visited in sequential order because of back mutations and conversions.

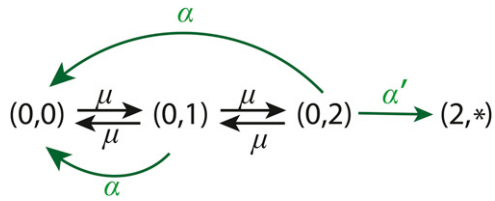

Figure 5 Markov chain for a evolving population in the absence of the direct path ($\delta = 1$) in the monoclonal approximation. Black arrows denote ordinary mutations, and green arrows denote gene conversions.

The fixation probability of a mutation from a state with fitness F_i to a state with fitness F_j is $p(s_{ij})$ from Equation (4) (diffusion approximation), where $s_{ij} = (F_j - F_i)/F_i$ is the fitness advantage of state j over state i [this approximation is known to break down for large selective advantages $s \gtrsim 0.1$ (Otto and Whitlock 2006). A branching process approximation (Uecker and Hermisson 2011; Bittihn *et al.* 2017) could be used instead for strongly beneficial mutations]. Multiplying this fixation probability by the corresponding mutation or conversion rates, we obtain transition rates between the different monoclonal states of the system. For the gene conversion path, the neutral mutations between states $(0,0)$, $(0,1)$ and $(0,2)$ occur with rate μN and become fixed with probability $1/N$, leading to the transition rate μ. Similarly, neutral conversions from (0,1) and (0,2) back to (0,0) have transition probability α. The transition rate for the conversion from state (0,2) to state (2,2) is $\alpha' = \alpha N p(s)$. The subsequent transitions among states (2,2), (2,1), and (2,0) are not important since we are only interested in fixation of the active gene, thus we can lump these three states into one state (2,*).

Thus, the evolution along the gene conversion path can be reduced to a four-state Markov chain with the single absorbing state (2,*), see Figure 5. The average time to reach state (2,2) from state (0,0) can be computed as

$$T_2 = \frac{\alpha^2 + \alpha'\alpha + 4\mu\alpha + 3\mu^2 + 3\alpha'\mu}{\alpha'\mu^2} \tag{16}$$

(see File S1). Numerical simulations of the full stochastic model agree with Equation (16) for small enough population sizes (Figure 6A). The time to fixation in a high-fitness state (2,2) is again large for both small and large α and has a minimum at an intermediate $\alpha_m = \mu\sqrt{3/[1 + Np(s)]}$.

For the direct path, a corresponding monoclonal approximation can be derived under the conditions mentioned at the

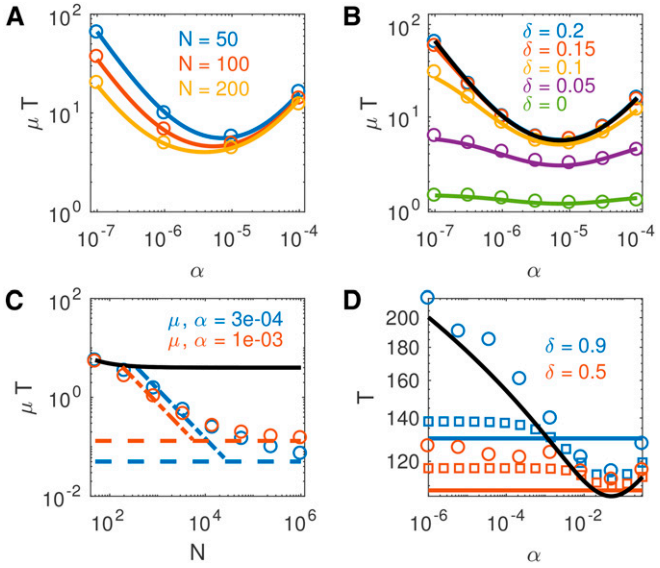

Figure 6 Simulations of the full stochastic model (Figure 1B) with $s = 0.1$. Fixation times T obtained as averages from 5000 independent simulations unless stated otherwise. Symbols represent numerical simulations. (A) Fixation times for $\mu = 10^{-5}$ and $\delta = 1$ as a function of α for different N. Lines represent the monoclonal approximation, (16). (B) Fixation times for $\mu = 10^{-5}$ and $N = 50$ compared to the monoclonal approximation for the gene conversion path (16) (black line), for different values of δ. The colored lines are analytical predictions combining the monoclonal approximation for the direct and the conversion path according to $T = (T_1^{-1} + T_2^{-1})^{-1}$. (C) Fixation times for $\delta = 1$ as a function of N for two different combinations of α and μ, compared to the monoclonal approximation (16) (solid line), the tunneling approximation (dashed-dotted lines) and the deterministic approximation (11) (horizontal dashed lines). Averages for larger N were obtained from progressively fewer numerical simulations. (D) Fixation times for $\mu = 10^{-3}$, large population size $N = 10^6$ and two different δ from 50 simulations (circles) compared to deterministic simulations (squares) and Equations (10) (color lines) and (11) (black line).

beginning of this section. The fixation time for the beneficial mutation without gene conversion in this regime is given by

$$T_1 = \frac{p(-\delta) + (1-\delta)\left[p(\delta) + p(\delta+s)\right]}{N(1-\delta)\mu p(-\delta)p(\delta+s)}. \qquad (17)$$

Again, if the direct path and the conversion path are assumed to be independent, the two approximations can be combined to yield an overall fixation time of $T = (T_1^{-1} + T_2^{-1})^{-1}$. A comparison with numerical simulations for different depths of the fitness valley is shown in Figure 6B. For these parameters, both the gene conversion path and the direct path are in the monoclonal regime. For large enough δ, the fixation time stays close to that predicted for the gene conversion path alone, as the direct path is strongly suppressed for deep fitness valleys.

It is interesting to note that substituting the optimal α_m into Equation (16) yields a minimum achievable T_2 for the gene conversion path that is proportional to $1/\mu$. Since T_1 follows the same scaling law, the speed-up factor provided by gene conversion with an optimal rate is independent of μ as long as both paths fulfill the conditions for the applicability of the monoclonal approximation, which is in contrast to the behavior of the stochastic tunneling approximations.

Behavior across different regimes

Figure 6C shows the behavior of the fixation time along the gene conversion path for increasing population size N and otherwise fixed parameters. For very small N, the monoclonal approximation for the fixation time applies (black line). The stochastic tunneling approximation (dashed-dotted lines) correctly predicts the initial behavior when the population size leaves the monoclonal regime. For larger population sizes, the numerical results then start deviating from this prediction before eventually converging to the deterministic prediction (dashed lines).

The fixation time in the deterministic theory is independent of the population size, as long as it is large enough. However, when the population size drops below the applicability of the deterministic predictions, the actual fixation time may strongly deviate from these predictions. Figure 6D shows a comparison of the deterministic theory to deterministic and stochastic simulations for a parameter combination that is on the border of the deterministic regime. For relatively shallow fitness valley $\delta = 0.5$ (red), both deterministic and stochastic simulations are in agreement with deterministic theory: for small α (where $T_1 < T_2$), the data follows T_1 estimate, and, for large α (where $T_2 < T_1$), it follows T_2 estimate. But for $\delta = 0.9$ (blue), stochastic simulations in the range of small α follow the deterministic estimate for the gene conversion path T_2 even though the deterministic prediction of T_1 in this range is smaller. The explanation of this discrepancy is that for $\delta = 0.9$ (blue), the direct path violates the condition $N \gg \delta/[\mu^2(1-\delta)]$, which leads to a severe underestimation of the corresponding fixation time T_1 by the deterministic theory. However, for the same parameters, the deterministic theory still works for the gene conversion path, and the corresponding T_2 is smaller than the actual T_1 for the direct path; therefore, full simulations follow the deterministic gene conversion path prediction.

A diagram summarizing different regimes of the gene conversion path and their population size requirements can be found in Figure 7A for a fixed choice of the mutation rate μ and the selective advantage s of the final mutants [an analogous diagram delineating similar regimes for the direct path was constructed by Weissman *et al.* (2009), so we do not repeat it here]. For increasing population size, the dynamics switch from monoclonal to stochastic tunneling, and then to deterministic. The gaps between these "pure" regimes correspond to the mixed regimes where not all transitions between states along the gene conversion path can be treated in the same manner: between the monoclonal and the tunneling regime, the population can get fixed in some states but tunnels through others. Similarly, some transitions may behave deterministically while others are still in the tunneling regime (between tunneling and deterministic).

Figure 7B compares the potential speed-up of fixation by gene conversion across the different regimes (T_1 and T_2 separately can be found in Figure S3 in File S1). As can be seen

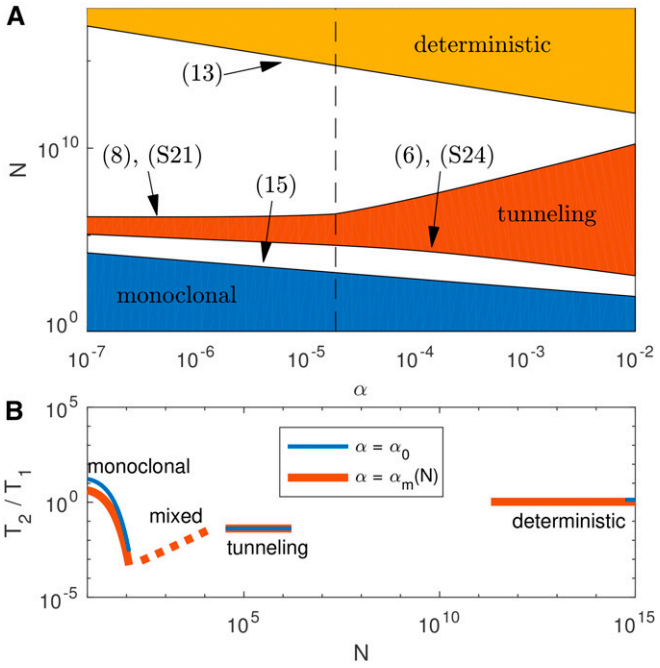

Figure 7 Summary of different approximations. (A) Regimes of validity for the different approximations of the gene conversion path in the minimal model in terms of the population size N and the conversion rate α for fixed $\mu = 10^{-6}$ and $s = 0.1$. The vertical dashed line indicates $\alpha = 2^{-4/3}(s\mu^2)^{1/3}$. (B) Speed-up T_2/T_1 provided by gene conversion across different population sizes for fixed $\mu = 10^{-6}$, $s = 0.1$ and $\delta = 0.1$ (the individual fixation times of the two paths can be found in Figure S3 in File S1). Results for each regime are plotted where the population size requirements for both the direct and the gene conversion path are fulfilled and both paths are in the same regime, except for the dashed red line labeled "mixed," where the direct path is in the tunneling regime and the gene conversion path is in the monoclonal regime (cf. Figure S3 in File S1). Calculations are shown for a fixed gene conversion rate $\alpha_0 = (16 \cdot 10^{13})^{-1/3}$ (the optimum in the tunneling regime; blue lines) and the optimal α for each regime and population size (red lines).

directly from the corresponding approximations in the previous sections, T_2/T_1 does not depend on N in the deterministic and the tunneling regime, but the numerical value of T_2/T_1 is much smaller in the stochastic tunneling regime. For the parameters used here, the gene conversion path is actually slower than the direct path in the deterministic regime, even for an optimal gene conversion rate α. Significant speed-up in this regime can only be achieved if δ is very close to 1 when the direct path is effectively blocked (*cf.* Figure 4). But, for smaller populations, stochastic effects make the direct path much less likely to be used even for modest δ, and so they benefit much more from the availability of the alternative gene conversion route. For very small population sizes (in the monoclonal regime) T_2/T_1 increases again, *i.e.*, the advantage of gene conversion is lost. This is intuitive, as, at small δN, the intermediate state becomes effectively neutral, and the fixation of the intermediate mutant in the active gene becomes more likely (see fixation time of the direct path only, Figure S3A in File S1). In this case, gene conversion will have a disadvantage because of its longer path length.

Rugged fitness landscapes

In the previous sections, we studied evolution in a very simple nonmonotonic fitness "landscape" consisting of just three states. However, it is not clear whether gene conversion could provide a similar speedup of adaptation in more complex fitness landscapes due to their high dimensionality. To explore this possibility, we now explicitly model the evolution of an $\mathcal{N}$-bit string representing a DNA sequence, building on the widely known NK model of epistatic interactions (Kauffman 1993). The fitness landscape maps each of $2^{\mathcal{N}}$ possible bit strings to a fitness value f between 0 and 1. NK landscapes derive their ruggedness from epistatic interactions of each bit with $\mathcal{K}$ other bits, which determine its fitness contribution. The ruggedness increases with $\mathcal{K}$ and leads to the formation of fitness peaks (Østman *et al.* 2010). Overall, the structure of the fitness landscapes produced by the NK model has been shown to be generally consistent with observations of pleiotropy and other genetic interactions in experimentally measured fitness landscapes (Wagner *et al.* 2008; Costanzo *et al.* 2010; Østman *et al.* 2012). We modified the classical NK model somewhat to account for the fact that, in reality, certain aspects of a genetic sequence have to be strongly conserved, *e.g.*, because they fulfill essential structural or catalytic functions. We therefore assume that $\mathcal{N}_0$ of the $\mathcal{N}$ bits are *essential*, and that the fitness f of a particular genotype is zero if any of these essential bits are altered.

As in our minimal model, we assume that the "gene" has undergone duplication, and consider an extended genotype augmenting the active $\mathcal{N}$-bits gene by an additional "passive region" of equal length $\mathcal{N}$ that does not impact fitness. Mutations correspond to the flipping of a single bit, while gene conversion is implemented by replacing the entire passive region with the sequence in the active region or vice versa. The dynamics of evolution and adaptation is then determined by the characteristic mutation rate per bit per individual μ and the gene conversion rate α per individual. We assume here that the population size N is small enough such that populations are monoclonal. Furthermore, we assume strong selection, such that beneficial mutations and conversions always lead to fixation, while deleterious mutations and conversions are always rejected. For neutral mutations like those in the passive region, the fixation probability is again $1/N$. These two simplifications—monoclonal population and strong selection—allow us to efficiently obtain statistics over many possible paths in several randomly generated fitness landscapes, and to observe the impact of gene conversion, since evolution away from a genotype without gene conversion is prohibited once it has reached a fitness peak.

We used the Gillespie algorithm (Gillespie 1977) to simulate evolution of a monoclonal population in this landscape after placing it in a random fitness peak and trace its fitness as it evolves over time. Figure 8A shows a sample trajectory of the actual fitness for a population initially trapped in a local fitness peak (thick solid black line) and the corresponding

potential fitness of passive region (thin solid green line). The symbols mark mutations and gene conversion events that change the potential fitness of the passive region. This example includes a single fitness improvement step at $t \approx 5.15$ due to gene conversion from the passive region to the active one, after a beneficial mutation is discovered as a result of two consecutive mutations in the passive region. Note that, immediately after this conversion, the actual fitness instantly climbs slightly above the potential fitness because the active region continues to evolve to the nearest fitness peak. While the actual fitness of individual trajectories is discontinuous, the ensemble-averaged fitness increases continuously and nearly linearly at small t (Figure 8, B and C). Therefore, we can use its initial slope r (the adaptation rate) as a quantitative measure for the advantage that gene conversion confers on the adaptability of the population.

Trade-off between mutations and conversions leads to optimal adaptation

Figure 8D shows that, as for the minimal model, the rate of fitness improvement also has a maximum at some intermediate value of α. For large α, the passive copy rarely evolves >1 bit-flip away from the genotype in the active region, because it is reset very frequently, and so the adaptation rate r decreases toward larger α (regime labeled r_1). For smaller α (regime labeled r_{2a}), the passive region has a certain fixed probability of discovering a beneficial mutation until at least one of the essential bits has mutated. A new round of discovery is started by each copying event from the active to the passive region, and so the discovery rate decreases when α, and, therefore the number of reset events, is reduced. In between these two extremes, there is a certain optimal rate of gene conversion that maximizes the adaptation rate r. It represents the trade-off between resetting the passive region frequently enough via gene conversion from active to passive, and, at the same time, allowing enough mutations to occur in between those events to discover a fitter genotype.

A more quantitative understanding of the adaptation by gene conversion in this landscape can be gained by considering the rates of three key processes: first, the rate with which neutral mutations emerge and get fixed in the passive region is $\mathcal{N}\mu$, since the mutation rate in the whole population is $\mathcal{N}N\mu$ and the fixation probability is $1/N$. Second, gene conversions from active to passive are also neutral, and therefore they get fixed with rate α. The third key process is the one which actually leads to fitness gains: gene conversion that transfers a beneficial mutation from the passive to the active region. The rate of this gene conversion is $N\alpha$, but its fixation depends on the state of the passive copy, *i.e.*, the probability P_b that the potential fitness of the passive region is larger than the actual fitness of the population at the time of conversion. If so, the active gene will quickly evolve to a new fitness maximum with a fitness increase Δf whose expectation value $\overline{\Delta f}$ can be measured numerically from a statistical analysis of our landscapes (see Figure S4D in File S1). If the potential fitness is less than the actual fitness, the passive-active conversion event is discarded.

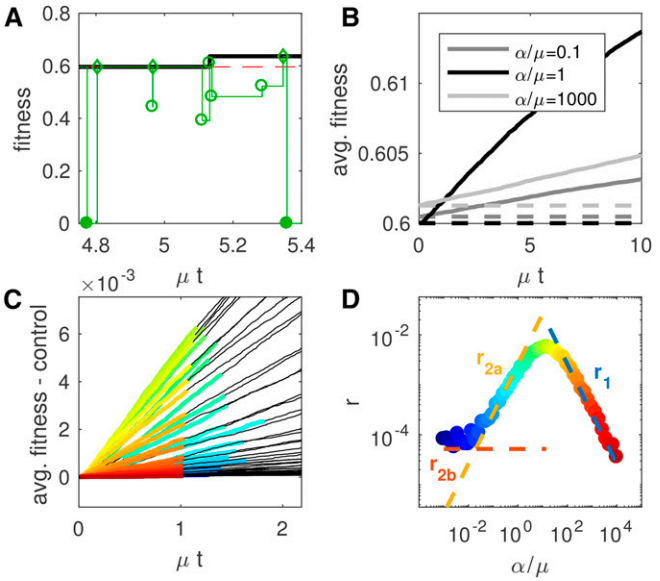

Figure 8 Fitness evolution in a rugged landscape. (A) Example trajectory of fitness of a single monoclonal population in landscape with $\mathcal{N} = 20$, $\mathcal{K} = 19$, $\mathcal{N}_0 = 10$ and $N = 10^5$ and $\alpha/\mu = 10$. Thick black line indicates actual fitness of the population, thin green line shows the potential fitness of the passive region a simulation with gene conversion, and red dashed line indicates the actual fitness in a simulation without gene conversion. Open circles denote the potential fitness immediately after mutations in nonessential bits of the passive region, solid circles denote the potential fitness (= 0) after mutations in essential bits, and diamonds denote potential fitness after gene conversion from active to passive region. (B) Ensemble-averaged fitness over time for select conversion rates α (solid lines) and control simulations without gene conversion (dashed lines). Other parameters are identical to those in (A). Fitness values were averaged over 1000 populations and 20 landscapes. (C) Initial average fitness trajectories for a range of different α. Colored sections indicate the intervals which were used to detect the initial rate of fitness increase r. The color corresponds to the value of α, *cf.* (D). (D) Initial rate of fitness increase r extracted from the data in (C). Dashed lines indicate the approximations of Equations (18) (blue), (19) (yellow), and (20) (red).

As long as $N\alpha \gg \mu\mathcal{N}$ (which might be the case even if $\alpha < \mu\mathcal{N}$ for sufficiently large population size N), conversion events are so frequent that potentially beneficial genotypes in the passive region are instantly copied to the active region and get fixed in the population (the passive region is "continuously activated" on the time scale of mutations). Therefore, gene conversion from passive to active is not rate limiting and we can ignore this step.

We can differentiate the following regimes (see File S1 for detailed derivations): (1) For large $\alpha \gg \mu\mathcal{N}$, the limiting factor is the ability of the passive region to accumulate k mutations before it is reset again by conversion from the active to the passive region. We know that $k > 1$ is required since the population always resides at a local maximum in the landscape. For small $\mu \ll 1$, the probability of finding a beneficial mutation at $k = 2$ is much greater than for any $k > 2$ so the latter can be neglected, leading to an adaptation rate of

$$r_1 \approx \overline{\Delta f}\frac{(\mu\mathcal{N})^2}{\alpha}P_2 \tag{18}$$

(we use the subscript i in r_i to identify the approximations of the adaptation rate r in different regimes). Here, P_2 is the fraction of genotypes at at Hamming distance $k = 2$ from the local maximum which are beneficial (see Figure S4A in File S1). (2) In the opposite limit $\alpha \ll \mu\mathcal{N}$, beneficial mutations can be discovered in two ways: (2a) in the wake of a conversion event from active to passive while the essential bits have not yet mutated; and (2b) by randomly discovering a beneficial genotype anywhere in the fitness space long after the conversion event from active to passive, when the initially correct configuration of the essential bits has already been lost. The latter is highly unlikely since it requires that all $\mathcal{N}_0$ essential bits are again correct by chance, *i.e.*, are in one out of $2^{\mathcal{N}_0}$ configurations. Therefore, the contribution of (2b) can only become comparable to that of (2a) for very small α, when the exploration phase of the passive region is very long to balance this exponentially small probability. Because of the ruggedness of the landscape, we assume that altering any bit (other than the essential bits) has the same fixed probability P_g of yielding a fitter genotype. This probability is equal to the fraction of all genotypes with correct essential bits that have a higher fitness than the current maximum (see Figure S4A in File S1), leading to

$$r_{2a} = \overline{\Delta f}\frac{\mathcal{N}-\mathcal{N}_0}{\mathcal{N}_0}P_g\alpha \tag{19}$$

for case (2a). While r_1 decreases with α as $1/\alpha$, r_{2a} is a linear function of α, thus an optimum speedup indeed occurs at the intermediate conversion rate $\alpha \propto \mu\mathcal{N}$. In case (2b), *i.e.*, for random discovery of beneficial genotypes (which are still continuously activated), the rate of fitness increase is

$$r_{2b} = \overline{\Delta f}\mu\mathcal{N}2^{-\mathcal{N}_0}P_g \tag{20}$$

Note that r_{2b} does not depend on α since the generation of new genotypes in the passive region happens with rate $\mu\mathcal{N}$ and any beneficial passive genotype is reliably activated.

The estimates (18), (19), and (20) are shown in Figure 8D superimposed with our numerical results. Figure S5A in File S1 in addition shows r as a function of the gene conversion rate α for different values of the ruggedness parameter $\mathcal{K}$. As can be seen, the approximations (18), (19), and (20) hold remarkably well across all levels of ruggedness.

For even smaller $\alpha \ll \mu\mathcal{N}/N$ (regime 3), the passive copy is no longer "continuously activated" (in contrast to all other cases considered above), and, therefore, many mutations occur in the passive region before the next gene conversion event from the passive to the active region. A very slow fitness gain can also be expected in this regime as a result of random discovery of beneficial genotypes anywhere in the fitness space, with the average speedup rate given by

$$r_3 \approx \overline{\Delta f}N\alpha 2^{-\mathcal{N}_0}P_g. \tag{21}$$

(see File S1 and Figure S5B in File S1). We should note, however, that, in reality, gene conversion itself might become less likely or cease completely when sequences of the original and the copy diverge sufficiently far (Teshima and Innan 2004). In light of this, the regimes of gene conversion via random discovery (regimes 2b and 3) would likely be suppressed. However, the optimal speedup occurs in the large α regime in which the homology between the original and the copy is well preserved, and our assumption of fixed conversion rate is justified.

Discussion

The idea that gene duplication might facilitate evolution has been around since the 1970s, when Ohno (2013) suggested that duplicate DNA sequences may escape strong selective pressure that usually prevents innovation beyond small incremental steps. It has also been argued that gene duplication and subsequent deactivation, mutation, and reactivation might have been an important mechanism during early evolution (Koch 1972), but consensus emerged that the most probable fate of the new copy is pseudogenization (Lynch and Conery 2000; Innan and Kondrashov 2010). Concerted evolution due to gene conversion only recently entered the picture; however, it has already been suggested as a possible mechanism of accelerated evolution of the *Nspb* gene family in *Caenorhabditis elegans* (Thomas 2006). A theoretical study by Mano and Innan (2008) explaining this observation investigated the fixation of a single mutation within multiple *active* copies of a gene subject to gene conversion. Previous studies implicated interlocus gene conversion from nonfunctional to functional genes as an important molecular mechanism leading to inherited diseases (Chen *et al.* 2007), but the evolutionary implications of this process have not been thoroughly addressed.

Our study provides the first theoretical description of adaptive evolution of a functional gene on a rugged fitness landscape in the presence of gene conversion with its nonfunctional duplicate. We demonstrate that gene conversion can significantly speed up adaptation: as long as there is a homologous region in the duplicated sequence that can serve as the template and target of gene conversion events, it can help to explore paths of adaptation in nonmonotonous landscapes that would otherwise be dramatically slowed down due to deleterious intermediate steps.

As the comparison across the different regimes in our minimal model shows, gene conversion provides a particular advantage for smaller populations due to stochastic finite size effects that render direct crossing of even shallow fitness valleys unlikely, while deterministically gene conversion only leads to a moderate speed-up for extremely deep fitness valleys. In the example in Figure 7B, the maximum speedup occurs at a population size on the order of 100 individuals. Because of this general trend, one might therefore speculate that the maintenance of pseudogenes for the

purpose of optimizing adaptation is more beneficial for higher organisms with relatively small population sizes, as opposed to, *e.g.*, bacteria. It is important to bear in mind though that the advantage of gene conversion is lost when the selective disadvantage of the intermediate mutants gets very small (equivalently, the population size may become so small that the fitness valley is effectively flat). In this case, adaptation cannot benefit from the existence of the gene conversion path as the additional mutation required slows its exploration.

In our modeling, we did not consider the process of gene duplication that originally led to the creation of the copy, neither did we consider the possibility of elimination of the duplicate gene. Consequently, we did not specify how the passive copy was freed from selective pressure, but since the "gene" we consider might actually be only a part of a larger, functional gene, there are a variety of possible mechanisms which would render the copy inactive and still leave a large enough region intact for gene conversion. These mechanisms include pseudogenization (of the larger gene) via premature stop codons, heterochromatin silencing, or mutations in *cis* regulatory regions of the copy. Our results are general in that they do not depend on this mechanism.

Across all dynamical regimes in the minimal model as well as in the rugged landscape model, we find an optimal rate of gene conversion maximizing the adaptation rate of the original gene (note, however, that this optimum does not always lead to an improvement over ordinary adaptation of the original gene, *cf*. Figure 7B). In both cases, this optimum is due to a trade-off: on the one hand, the passive copy must be able to explore the genetic space, which is only possible for small gene conversion rates α, when the passive copy is not reset to wild type too frequently. On the other hand, a beneficial mutation must be present and then transferred to the active region, which requires a sufficiently large α.

The first requirement is qualitatively similar in both our models: for too large gene conversion rates α, the passive gene gets stuck in the state of the active gene. However, the second requirement (the reason why too small gene conversion rates are also suboptimal) is due to different reasons in our two models: while in the NK-landscape model, the accumulation of deleterious mutations prohibits finding beneficial genotypes at small gene conversion rates (as only the reset to the active version is likely to reverse all deleterious mutations in the passive region), the minimal model only considers two loci and ignores other mutations. In the latter, a large enough gene conversion rate is required simply because the beneficial mutation needs to be transferred to the active region. This is usually not a concern in the NK-landscape model: if the gene conversion rate α is on the order of the total mutation rate of any bit $\mu\mathcal{N}$ [the region of the optimum in between approximations (18) and (19)], this automatically implies $N\alpha \gg \mu\mathcal{N}$ for sufficiently large populations leading to reliable transfer to the active region. Even though the fixation of beneficial mutations in the passive region might be rare without the constraint of natural selection, their reactivation in successful cases would be very likely. Thus, together, our two models point to two potential reasons why a sufficiently large the gene conversion rate is required for efficient adaptation.

Note also that we assumed equal conversion rates α in both directions, whereas, in reality, the underlying molecular mechanisms and the genetic context could potentially lead to differences between the forward and the backward rates. In addition, gene duplication itself may also contribute to the forward rate of resetting the passive region to the original genotype, as it produces a fresh copy that can get silenced and become a new template for gene conversion as described by our model.

The quantitative applicability of our findings to experimental evolution is limited by our choice of model fitness landscapes. Natural fitness landscapes may have quite different structure than abstract NK-type landscapes, which could affect our estimates. For example, if finding a beneficial genotype required $k > 2$ single mutations, the adaptation rate r_1 (in the regime where resetting of the silent copy is rate limiting) would be proportional to $(\mu\mathcal{N})^k$; however, the optimal conversion rate would still scale as $\mu\mathcal{N}$. Our theory also assumed that the probability of yielding a fitter genotype, which we used to obtain Equation (19), is the same for any mutation and any local fitness peak. In reality, the fitness landscape parameters may vary greatly. Nonetheless, we believe that the salient qualitative features of evolutionary dynamics accelerated by gene conversion, including the presence of an optimal conversion rate commensurate with the mutation rate, are robust, and can be expected to play an important role in adaptive evolution.

Data availability

The authors state that all information necessary for confirming the conclusions presented in the article is represented fully within the article and the Supplemental Material.

Acknowledgments

We are grateful to Michael McLaren for his invaluable comments on the manuscript, and to Anne Carvunis and Sergey Kryazhimskiy for stimulating discussions. This work was supported by National Science foundation (NSF) grant MCB-1616997 and the San Diego Center for Systems Biology, National Institutes of Health/National Institute of General Medical Sciences (NIH/NIGMS) grant P50-GM085764. P.B. acknowledges support from Human Frontier Science Program (HFSP) fellowship LT000840/2014-C.

Literature Cited

Bittihn, P., J. Hasty, and L. S. Tsimring, 2017 Suppression of beneficial mutations in dynamic microbial populations. Phys. Rev. Lett. 118: 028102.

Chen, J.-M., D. N. Cooper, N. Chuzhanova, C. Férec, and G. P. Patrinos, 2007 Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. 8: 762–775.

Costanzo, M., A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear *et al.*, 2010 The genetic landscape of a cell. Science 327: 425–431.
Elder, J. F. Jr., and B. J. Turner, 1995 Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. 70: 297–320.
Fawcett, J., and H. Innan, 2011 Neutral and non-neutral evolution of duplicated genes with gene conversion. Genes (Basel) 2: 191–209.
Fisher, D. S., 2007 Course 11 "evolutionary dynamics", pp. 395–446 in *Complex Systems. Lecture Notes of the Les Houches Summer School*, Vol. LXXXV, edited by J. P., Bouchaud, M. Mezard, and J. Dalibard. Springer-Verlag, Amsterdam.
Flyvbjerg, H., and B. Lautrup, 1992 Evolution in a rugged fitness landscape. Phys. Rev. A 46: 6714–6723.
Gillespie, D. T., 1977 Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81: 2340–2361.
Hastings, P. J., 2010 Mechanisms of ectopic gene conversion. Genes (Basel) 1: 427–439.
Hayakawa, T., T. Angata, A. L. Lewis, T. S. Mikkelsen, N. M. Varki *et al.*, 2005 A human-specific gene in microglia. Science 309: 1693.
Innan, H., 2009 Population genetic models of duplicated genes. Genetica 137: 19–37.
Innan, H., and F. Kondrashov, 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11: 97–108.
Innan, H., and W. Stephan, 2001 Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. Genetics 159: 389–399.
Iwasa, Y., F. Michor, and M. A. Nowak, 2004 Stochastic tunnels in evolutionary dynamics. Genetics 166: 1571–1579.
Kauffman, S. A., 1993 *The Origins of Order: Self Organization and Selection in Evolution*. Oxford University Press, New York.
Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics 47: 713–719.
Koch, A. L., 1972 Enzyme evolution: I. The importance of untranslatable intermediates. Genetics 72: 297–316.
Kondrashov, F. A., and E. V. Koonin, 2004 A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. Trends Genet. 20: 287–290.
Liao, D., 1999 Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. 64: 24–30.
Lin, Y.-S., J. K. Byrnes, J.-K. Hwang, and W.-H. Li, 2006 Codon-usage bias *vs.* gene conversion in the evolution of yeast duplicate genes. Proc. Natl. Acad. Sci. USA 103: 14412–14416.
Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.
Mano, S., and H. Innan, 2008 The evolutionary rate of duplicated genes under concerted evolution. Genetics 180: 493–505.
Ohno, S., 2013 *Evolution by Gene Duplication*. Springer Science & Business Media, Berlin.
Østman, B., A. Hintze, and C. Adami, 2010 Critical properties of complex fitness landscapes. arXiv:1006.2908.
Østman, B., A. Hintze, and C. Adami, 2012 Impact of epistasis and pleiotropy on evolutionary adaptation. Proc. Biol. Sci. 279: 247–256.
Otto, S. P., and M. C. Whitlock, 2006 *Fixation Probabilities and Times*. John Wiley & Sons, Ltd., Hoboken, NJ.
Paulsson, J., M. El Karoui, M. Lindell, and D. Hughes, 2017 The processive kinetics of gene conversion in bacteria. Mol. Microbiol. 104: 752–760.
Plata, G., and D. Vitkup, 2014 Genetic robustness and functional evolution of gene duplicates. Nucleic Acids Res. 42: 2405–2414.
Saakian, D. B., A. S. Bratus, and C.-K. Hu, 2017 Crossing fitness canyons by a finite population. Phys. Rev. E 95: 062405.
Serra, M. C., and P. Haccou, 2007 Dynamics of escape mutants. Theor. Popul. Biol. 72: 167–178.
Takuno, S., T. Nishio, Y. Satta, and H. Innan, 2008 Preservation of a pseudogene by gene conversion and diversifying selection. Genetics 180: 517–531.
Teshima, K. M., and H. Innan, 2004 The effect of gene conversion on the divergence between duplicated genes. Genetics 166: 1553–1560.
Thomas, J. H., 2006 Concerted evolution of two novel protein families in Caenorhabditis species. Genetics 172: 2269–2281.
Uecker, H., and J. Hermisson, 2011 On the fixation process of a beneficial mutation in a variable environment. Genetics 188: 915–930.
Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman *et al.*, 2008 Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452: 470–472.
Walsh, B., 2003 Population-genetic models of the fates of duplicate genes. Genetica 118: 279–294.
Weinreich, D. M., and L. Chao, 2005 Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59: 1175–1182.
Weissman, D. B., M. M. Desai, D. S. Fisher, and M. W. Feldman, 2009 The rate at which asexual populations cross fitness valleys. Theor. Popul. Biol. 75: 286–300.
Wu, N. C., L. Dai, C. A. Olson, J. O. Lloyd-Smith, R. Sun *et al.*, 2016 Adaptation in protein fitness landscapes is facilitated by indirect paths. Elife 5: e16965.

Communicating editor: J. Masel