Breaking evolutionary constraint with a trade-off ratchet

Marjon GJ de Vos [a,b,c], Alexandre Dawid [d], Vanda Sunderlikova [a], Sander J Tans [a*]

[a] FOM Institute AMOLF, Science Park 104, 1098 XG, Amsterdam, The Netherlands.

[b] IST Austria, Am campus 1, 3400, Klosterneuburg, Austria.

[c] Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

[d] Univ. Grenoble Alpes, LIPhy, 140 Avenue de la Physique, F-38402 Saint Martin d'Hères, France.
CNRS, LIPhy, 140 Avenue de la Physique, F-38402 Saint Martin d'Hères, France.

*Corresponding author: s.tans@amolf.nl

BIOLOGICAL SCIENCES: Evolution, Systems Biology, Biophysics and Computational Biology.

Abstract

Epistatic interactions can frustrate and shape evolutionary change. Indeed, phenotypes may fail to evolve because essential mutations are only accessible via positive selection if they are fixed simultaneously. How environmental variability affects such constraints is poorly understood. Here we studied genetic constraints in fixed and fluctuating environments, using the *Escherichia coli lac* operon as a model system for genotype-environment interactions. We found that in different fixed environments, all trajectories that were reconstructed by applying point mutations within the transcription factor-operator interface became trapped at sub-optima where no further improvements were possible. Paradoxically, repeated switching between these same environments allows unconstrained adaptation by continuous improvements. This evolutionary mode is explained by pervasive cross-environmental trade-offs that reposition the peaks in such a way that trapped genotypes can repeatedly climb ascending slopes, and hence escape adaptive stasis. Using a Markov approach, we developed a mathematical framework to quantify the landscape crossing rates and show that this ratchet-like adaptive mechanism is robust in a wide spectrum of fluctuating environments. Overall, this study shows that genetic constraints can be overcome by environmental fluctuations, and that cross-environmental trade-offs thus not necessarily impede, but can also facilitate adaptive evolution. As trade-offs and environmental variability are ubiquitous in nature, we speculate this evolutionary mode to be of general relevance.

Significance

Sub-optimal fitness peaks are generally recognized as causing evolutionary stasis. Here, we show that these constraints can be overcome in an adaptive manner, by reconstructing mutational trajectories for a transcription factor and its DNA binding site in variable environments. Cross-environmental trade-offs, typically associated with evolutionary limitations, are an essential enabling component of this evolutionary mechanism. Our results underscore the importance of characterizing environmental dependencies when studying genetic interactions, and provide the clearest indication so far that environmental variability can accelerate evolution hampered by stasis in constant conditions. Given that environmental variations and trade-offs are ubiquitous, this evolutionary mechanism may be relevant to a wide range of genetically constrained phenotypes and major evolutionary transitions.

Introduction

It is widely believed that epistatic interactions can direct evolutionary change (1, 2) (3-7). Epistasis has been implicated in shaping RNA (8) and protein (4, 6, 7, 9) sequences, sensing (5) and translation (10) functions, as well as developmental programs (11) and speciation (12-14). Phenotypes may be difficult to evolve not because they are impossible biochemically or physically, but because essential mutations are mutually dependent, and must be fixed together to be selected positively (5, 15-17). How such genetic constraints can be overcome has been considered previously: population expansion or subdivision can limit negative selection and maintain less fit phenotypes (18, 19), large populations and long waiting times can enable the joint fixation of multiple mutations (20), while recombination can join mutant alleles (21-23). Other mechanisms include drift (24-26), partial penetrance (27) and non-heritable life time plasticity (28, 29). However, how the constraining effects of such genetic interactions are affected by environmental variability remains poorly understood. It has been shown that mutational effects (30-33) and epistasis itself (34, 35) can depend on the environment, that bacterial resistance evolution can be contingent on the rate of antibiotic increase (36), and that adaptation *in silico* can be accelerated by environmental change (37-40). These observations suggest that the effects of environmental variability may go beyond merely producing variable selective pressures that favour certain phenotypes, but could also be involved in controlling phenotype accessibility and stasis.

To investigate how environmental variability affects genetic constraints, we focused on a model system for genotype-environment interactions, the *lac* regulatory system of *E. coli*. Its physiology has been studied extensively: in the presence of lactose, expression of the *lac* genes allows *E. coli* cells to import and metabolize lactose, while in the absence of lactose, repression of these genes limits physiological costs (41, 42). The ability to regulate *lac* expression relies on the binding of the *lac* repressor to the *lac* operator DNA sequence upstream of the coding region (Fig. 1A). We surmised that the co-evolution of such protein-DNA interfaces could be severely constrained by epistatic interactions, such that some genotypes are inaccessible by positive selection in single mutation steps. In lock-key recognition, mutating either lock or key is expected to lead to recognition loss (2, 15). At the same time, mutating both lock and key may produce a different, better-matching pair. Indeed, the *lac* transcription factor phylogeny suggests extensive historic adaptation of the repressor-operator interface, and reveals multiple homologous repressors that bind specifically to their cognate operator (43, 44). Furthermore, the lock-key recognition of the *lac* regulatory system is a highly specific function that is confined to a limited number of residues, which restricts the range of adaptive solutions. Indeed, extensive mutational analysis of the *lac* repressor-operator interface has shown that just two repressor residues and four operator bases control binding specificity (Fig. 1A and B) (45, 46). Moreover,

in contrast to phenotypes that are affected by many unknown mutations, one can identify the genetic interactions between mutations in controlling residues that are key to genetic constraint.

Results

Trajectories are not accessible in constant environments

We mutated sites in the *lac* operator DNA and the multimeric *lac* repressor that control binding specificity (Fig. 1A and B), and measured the expression level of the downstream *lac* genes in two contrasting conditions (Methods). We quantified the ability to repress the *lac* genes, *R*, as the inverse of the measured *lac* expression level in the *absence* of inducing ligand (Fig. 1C, Methods). The ability to express the *lac* genes, *E*, was quantified by the measured *lac* expression level in the *presence* of ligand (Fig. 1C). Note that the repression ability (*R*) is thus not the inverse of the expression ability (*E*). First, we constructed four *lac* repressor-operator variants that have been predicted (45-48) to display binding: PK:agga, PS:acca, MK:acca, and YQ:tggt, where the first two letters indicate the controlling repressor residues and the last four the controlling operator bases (Fig. 1A and B). We focused on the latter two variants, as they displayed substantial fold-changes between the induced and non-induced expression level ($R \bullet E$ was 6 and 55 respectively), with *E* being approximately equal but *R* about 20-fold lower for MK:acca. The MK:acca genotype is thus able to regulate *lac* expression, but can improve repression ability by mutating the repressor (MK to YQ) and operator (acca to tggt).

We investigated the interaction between non-cognate pairs by swapping around the two operators. The ability to repress (*R*) was found to be low (i.e. expression in the absence of ligand was high) for MK:tggt and YQ:acca (100 to 200-fold lower than for the cognate pair YQ:tggt). These data were consistent with the reciprocal sign epistasis hypothesised for lock-key interactions: changing either of the binding partners alone leads to binding loss, but changing the other partner as well restores it. This notion was supported by the overall expression levels for MK:tggt and YQ:acca, which were high and unresponsive to ligand ($R \bullet E = 1$). However, while the presence of reciprocal sign epistasis is required, it is not sufficient to constrain phenotypes on sub-optima (15). Indeed, the repressor and operator modifications both involve multiple mutations, and their one-by-one fixation in particular order (1) could confer continuous improvements in repression ability.

To test the accessibility of trajectories considering all possible orders of all essential mutations, we constructed the remaining intermediate genotypes between MK:acca and YQ:tggt. In total 6! = 720 direct trajectories can be taken along the $2^6 = 64$ genotypes. Whether a mutation is positively selected and thus accessible depends on the sign, rather than the magnitude, of the associated fitness change. Fitness and

phenotypic changes have the same sign when the phenotype-fitness relation is monotonic. In the environment without ligand for instance, a trajectory is then not accessible when it contains mutations that decrease *R*. Below we first consider this case of monotonic phenotype fitness relations, and later relaxi this assumption to also consider a range of non-monotonic relations.

Analysis showed that all trajectories contain depressions in both *R* and *E* (Fig. 2, Fig. S1). The depressions are at least 2 mutations wide and peak at a width of 5 mutations, while the involved decrease is at least 3-fold and reaches up to approximately 100-fold. Thus none of the trajectories to YQ:tggt is accessible by fixing mutations one-by-one in either of the two environments. While this analysis concerns only direct trajectories, i.e. without mutational reversions, allowing for reversions did not open up accessible trajectories in either of the environments (Text S1-S3). Overall, these data indicate that higher-order genetic interactions (i.e. epistasis involving multiple mutations) limit optimization of the *lac* regulatory phenotype in each of the two environments.

Fluctuating environments allow gradual optimization

How does environmental variability affect these constraints? We first explored this question with individual trajectories starting with MK:acca. For instance, *R* can be increased through an operator mutation (MK:acca to MK:tcca) in the environment without ligand, but then remains trapped because the other mutations yield no further improvements (Fig. 3A and B, and Fig. S2). However, switching to the other environment opens up various trajectories that increase *E*, such as the repressor mutation MK:tcca to MQ:tcca. After a further increase in *E* (MQ:tcca to MQ:tgca), the system becomes trapped again on a suboptimum. Concomitantly, note the low *R* for this genotype (Fig. S2A) which indicates an overall weakened binding between these LacI and operator variants. Switching back to the first environment now allows escape by a compensatory mutation that provides access to YQ:tggt by re-establishing binding and increasing *R*. We found that a significant fraction of the direct trajectories (21%) becomes accessible in this manner (Fig. S1 and S3). Accessibility was afforded by diverse patterns of environmental change, though not by all (Table S1). Interestingly, the number of accessible trajectories starting in the environment *E* (with IPTG) outnumbered the accessible trajectories starting in the environment *R* (without IPTG) (92 vs 56), even though the main improvement is made in *R*. Overall, these findings indicate that mutational pathways that failed to confer gradual optimization in either constant condition can do so when alternating between these same conditions.

A ratchet mechanism that exploits trade-offs

The above findings cannot be understood from environmental interactions that alter the magnitude of mutational effects, as they would affect only the depth of constraining valleys. Rather, they indicate the importance of cross-environmental trade-offs, in which increases in *R* occur at the expense of decreases in *E*, and vice versa, increases in *E* lead to decreases in *R* (Fig. 3B). Such trade-offs are also referred to as GxE interactions (49), and can give rise to higher-order GxGxE (34) and sign environmental-epistasis (36). In the *lac* repressor-operator system, cross-environmental trade-offs between *E* and *R* were pervasive (55% of all mutational steps in the direct trajectories, Fig. S4), and can be understood mechanistically. For example, a low but significant level of repression can be maintained in the presence of inducer through residual binding (50). We found for several genotypes (22) that the induced expression level is significantly lower than the highest measured level for the involved operator (Fig. 1C); consistent with residual binding of induced repressors reducing expression. Mutations that increase (decrease) the overall repressor-operator affinity in both environments will increase (decrease) both the repression ability without inducer as well as the residual repression with inducer, leading to opposite effects on *R* and *E*, and hence to cross-environmental trade-offs (Fig. S4). These trade-offs have consequences for the relations between constraints in different conditions. We found multiple local optima for each of the two environments (3 in *R* and 13 in *E*), but none coincided at the same genotype. This feature allows trajectories to repeatedly surf ascending slopes and hence traverse valleys in a ratchet-like manner: when trapped on a local optimum, the system can wait for an environmental change that enables repositioning on a new ascending slope.

Crossing rates in fluctuating environments

To assess the robustness of this evolutionary mode for different environmental conditions we extended an evolution model based on a fixed-environment Markov process (51) to include environmental fluctuations (Fig. 4A-C, Methods, Text S1-S6 and Fig. S7 and S8). We considered a discrete-time Moran process in the strong selection weak mutation (SSWM) regime, in which mutational reversions are allowed and trajectories can be of arbitrary length (Methods) (52, 53). Here we first assume fitness and selection to be proportional to phenotype (*E* and *R*), and then test different non-linear and non-monotonic phenotype-fitness relations (Text S7, Fig. S9). Consistent with the observed constraint in fixed conditions (Fig. 2), we found that the rate to evolve to YQ:tggt from MK:acca (crossing rate k_c) is null for either constant environment (Fig. 4D). However, k_c is consistently above zero when the environmental fluctuation rate k_f is lower than the mutation rate k_m (Fig. 4D blue and green lines) and maximized when $k_f = k_m$, consistent with previous related work (39). This can be understood as follows: for $k_f >> k_m$ there is an effective averaging over the two environments resulting in a constrained condition, whereas for $k_f << k_m$ the waiting time for an environment-triggered escape is long enough to allow mutational escape in one of the two environments. Environmentally triggered escape is thus found to be robust to changes in the ratio between the times spent in the two environments (Fig. 4D).

The above result remained valid for non-linear phenotype-fitness relations, reflecting for instance expression costs that make the dependence weaker for higher *E* (Text S7, Fig. S9) (41, 54). The escape mechanism also remained robust for non-monotonic phenotype-fitness relations, in which fitness first increases with increasing *E* and then decreases. We found that the mechanism broke down when *E* increases do not confer fitness increases at all, reflecting a scenario in which *lac* expression costs outweighs the benefits in the inducing environment (Fig. S9). However, the latter is not observed experimentally (41, 42).

Computationally generated landscapes

To further probe the requirements for crossing multi-peaked landscapes, we considered computationally generated random landscapes. In constant conditions, mutational trajectories become trapped on these landscapes as expected, and hence the probability to find selectively accessible paths is low (Fig. S5). We found that when we progressively increase the level of cross-environmental trade-off (as quantified by the fraction of mutations having opposite effects in the two environments), the probability of finding accessible paths increases, until practically all of the landscapes that were generated contained accessible paths in fluctuating conditions at a trade-off level of 0.5 or more (Fig. S5). We note that this value was 0.47 for the experimentally determined landscape. Overall, these results indicated that cross-environmental trade-offs are a central ingredient for overcoming constraining genetic interactions in fluctuating environments by positive Darwinian evolution.

Discussion

To investigate how environmental variability affects genetic constraints within evolutionary trajectories, we have systematically mapped the genotype-phenotype landscape spanning two matching pairs of transcription factors and their DNA binding sites. Consistent with theoretical considerations on the evolution of molecular recognition, we find that the resulting landscapes in different contrasting environments are highly rugged: none of the mutational trajectories between the two pairs are selectively accessible by fixing one mutation at a time. Landscapes ruggedness has been studied previously for phenotypes that involve multiple possible genes and spontaneously evolved mutations (5, 16, 21, 55, 56). Here we studied a phenotype that has a well-understood genetic basis, is restricted to one physical site, and hence is controlled by a reduced set of mutations. The data show that epistatic interactions between these controlling mutations give rise to distinct peaks within genotype space.

Rugged landscapes constitute a specific type of genetic constraint that is distinct from, for instance, the presence of sign-epistasis, which imposes specific fixation order but does allow accessible paths (1).

Theoretical investigations have considered how escape from entrapment on suboptimal peaks is enabled; by limiting negative selection (18, 19), simultaneously fixing mutations (20), recombination (21-23), and drift (24-26). These mechanisms do not permit adaptive evolution by positive selection of single mutations. Here we find that such gradual adaptive evolution by positive selection is possible when considering temporal alternations between different environments, with cross-environmental trade-offs as a second essential ingredient. These trade-offs are pervasive in the quantified landscapes. They allow the environmental changes to displace local fitness peaks in such a way that the system can be ratcheted in an adaptive manner through fitness valleys by repeatedly climbing locally ascending slopes. We note that at high mutation rates, as for instance observed in viral evolution, multiple variants may arise before one sweeps through the population, resulting in clonal interference and competition between variants. While here the sweep time rather than the time between mutations is the relevant time scale, the environmental ratchet mechanism remains valid in this regime.

Crossing fitness valleys by positive selection may seem contradictory. However, one must distinguish between the global fitness of the population and the relative fitness differences between individuals. When trapped, the relative fitness changes conferred by mutational changes lead to negative selection within the population. In contrast, the global fitness decreases conferred by environmental changes affect the entire population and do not lead to negative selection. These global fitness decreases open up possibilities for mutations that confer relative fitness gains again. These repeated gains do not come completely for free, as the mechanism does not only drive forwards but also backwards: after a global fitness decrease induced by environmental change, positively selected mutations are opened up that point away from the ancestor but also back to it. However, this does not cancel the positive selection of this evolutionary mode, but merely reduces its efficacy. Thus, environmental changes can serve to overcome fitness decreases that constrain genotypes in constant environments.

The role of trade-offs in facilitating adaptive change is notable, as they are typically associated with adaptive constraint. Trade-offs can be rationalized mechanistically for regulatory phenotypes that rely on lock-key molecular recognition, which have been shown to be important genetic material for evolutionary novelties (57, 58). Evolutionary optimization in the presence of trade-offs can be seen as a sequence of compensatory mutations; recognition loss caused by mutating one binding partner can be restored by complementary mutations in the other binding partner. Note that trade-offs are present more generally in nature (59, 60), as are environmental fluctuations (61), and hence the reported environmental ratchet escape mechanism is not necessarily limited to regulatory phenotypes.

Organisms are known to occupy remarkable wide ranges of contrasting environments. On the one hand, our findings suggest that the regulatory systems that enable these wide environmental niches are constrained by pervasive (reciprocal) sign-epistatic interactions resulting in rugged landscapes. An inability to optimize regulatory responses to environmental change can result in stasis and the emergence of environmental specialists. However, we find that the proposed mode of adaptation can overcome these constraints, and facilitate the evolution of generalists that can occupy diverse niches. Interestingly, such adaptability comes at the cost of weakened environmental robustness of desirable genotypes. Indeed, the stabilizing selection of such a genotype at the top of a fitness peak can be disrupted by trade-offs and environmental change that together can displace peaks. Some environmental robustness may alternatively be afforded by regions in genotype space that are flat in both environments. Yet, except for a 'neutral chain' (YK:agga-YQ:agga-MQ:agga-MQ:tgga-MQ:tcga), no other neutral region was observed in our landscape.

Genetic constraints are commonly regarded as a key factor in adaptive stasis and major evolutionary transitions, such as for instance the evolution of sex, multicellularity or symbiosis, and eusociality (62, 63). Overcoming genetic constraints by trade-offs in a fluctuating environment could be relevant to these transitions, given its few requirements and their ubiquitous presence. Additionally, this evolutionary mode may have implications for clinically or biotechnologically relevant problems, such as the treatment of infections with multi-drug protocols (64-66) and the evolutionary engineering of antibodies (67). It will be of interest to test whether this adaptive mode affects the dynamics of experimental evolution in variable environments.

Methods

Strains

We used *Escherichia coli* strain BW23473(68) F-, Δ(argF-lac)169, ΔuidA3::pir+, recA1, rpoS396(Am), endA9(del-ins)::FRT, rph-1, hsdR514, rob-1, creC510 (Yale Coli Genetic Stock Collection #7837) for all experiments.

Plasmid

We used a single copy plasmid that can be induced to a multicopy plasmid for cloning purposes (pETcoco-2, Novagen), to mimic natural expression levels. *lacI* was deleted from the pETcoco-2, and replaced with the T0 and Trev terminators from the SacI-SpeI fragment of pInv-110 (69). The introduced placI was derived from the overexpressing $placI^{q}$ sequence from pTrc99A (70). By site-directed mutagenesis the t at -35 was replaced with the c nucleotide (Quick change site-directed mutagenisis kit, Qiagen), leading to native LacI expression levels. The auxiliary operator O3, located at the end of the coding sequence of the wild type *lacI*,

has a low affinity for LacI (about 300-fold lower than O1 (71)). It is involved in weak auto-repression of the *lac* repressor (72), and was hence deleted while preserving the LacI amino acid sequence (7). placI, combined with *lacI* from pRD007 (which lacks the auxiliary O3 operator sequence) (7) and a spacer sequence derived from Drosophila Kinesin (pCascade5) by overlap PCR (Pfu Ultra, Invitrogen) were inserted in pETcoco-2. The plac, *lacZ* and *lacY* wild-type sequences were derived from *E. coli* strain MG1655 (Yale Coli Genetic Stock Collection #6300). *lacA* was not incorporated in the assayed operon, as it decreases the active concentration of IPTG in the cell (73, 74). To facilitate the insertion of mutants (see below), the plac region was amplified until the 9th codon in the coding sequence of *lacZ* (Phusion polymerase, Finnzymes, New England Biolabs), a BmtI restriction site was introduced, while the wild type amino acid sequence was preserved. This fragment was inserted in the pETcoco-2 vector backbone. The remaining of the *lacZ* and *lacY* sequence was amplified (Phusion polymerase, Finnzymes, New England Biolabs) and inserted in the polylinker sequence of pETcoco-2 to make pMdV53. All restriction enzymes used and T4 ligase were obtained from New England Biolabs, all primers were obtained from MWG Eurofins.

Mutants

lacI and plac-operator mutants were constructed separately. plac-operator variants differed in the 4, 5, 5' and 4' nucleotide residues as depicted in Fig. 1B and C. This O1 operator was made palindromic by the deletion of the central c-g base pair (75). In LacI, amino acid residues 17 and 18 were altered, using optimal codons for the amino acid variants. Half-sides of each *lacI* or plac-operator variant were constructed using Pfu Ultra polymerase (Invitrogen) with one flanking primer and one primer harbouring the desired nucleotide variation. After gel purification (Gel extraction kit, Qiagen) and ExoI-treatment (New England Biolabs), the half-sides with the desired nucleotide variation in the region of overlap, were amplified by overlap PCR to obtain full length *lacI* or plac-operator variants. Combinations of *lacI*-spacer-plac mutants were constructed by overlap PCR (Pfu Ultra, Invitrogen), inserted in the pMdV53 backbone, and verified by Sanger sequencing.

Media

All growth and expression measurements were performed in MOPS EZ Rich Defined medium (Teknova, Hollister, CA, USA), with 0.2% glucose as carbon source, and supplemented with 1 mM thiamine-HCl, 1 mM uracil and appropriate antibiotics. For expression ability measurements the medium was supplemented with 1 mM IPTG.

Expression level measurements

Cultures were grown at 37°C in PerkinElmer Victor3 and Victor X3 plate readers with 200 µl of medium per well in a black clear-bottom 96 well plate (NUNC 165305). Optical density at 600 nm was recorded every 4 min, and every 29 min 9 µL of sterile water was added to each well to counteract evaporation. Between measurements the well plate was shaken at double orbit with a diameter of 2 mm. Cells were fixed after the cultures had reached an optical density in the plate reader of at least 0.015 and at most 0.07, by adding 20 µl FDG-fixation solution (109 µM fluorescein di-β-D-galactopyranoside (FDG, MarkerGene Technologies Inc), 0.15% formaldehyde, and 0.04% DMSO in deionized water). Fluorescence development was measured every 8 min (excitation 480 nm, emission 535 nm), as well as the optical density at 600 nm. Shaking and dispensing conditions were as described above. When cells were not induced with IPTG during growth, 1 mM IPTG was added to each well immediately before or after fixation as the assay is sensitive to the amount of IPTG present in the medium (7). Analysis of the fluorescence, to quantify the LacZ expression level, was as described in (7).

Statistical tests

For each genotype, expression measurements were performed in either triplicates or quadruplicates. Expression levels measured in one 96 well plate were normalised to a control genotype on that plate. Throughout all the analysis, whenever the values between two genotypes were compared (either expression or repression ability values, or relative fitness values obtained through the cost-benefit treatment), a two-tailed Welch's t-test was performed to test the null hypothesis that the two genotypes do not have significantly different values (hypothesis of neutrality). The null hypothesis was rejected if the *p-value* obtained from the test was lower than a significant level of 0.05. If the null hypothesis was not rejected, the mutations distinguishing the two genotypes were considered neutral, that is, e.g., we would take the condition $\overline{f_j} = \overline{f_i}$ in equation S1 of the Text S2. Else, the values would be considered to be significantly different and subsequent computation would be performed accordingly.

Model for the evolution dynamics in fluctuating environment

We modelled the dynamics of evolution in fluctuating environment using a discrete-time Moran process in the regime of strong selection weak mutation (SSWM) and through a heterogeneous Markov process.

Briefly, for each environment: without IPTG (environment *R*) or with IPTG (environment *E*), a 64×64 matrix of transition probabilities between the 2^6 = 64 isogenic states of the population is first computed based on the difference in performance between starting genotype g_i and the arriving genotypes g_j (see equations S3 and S4 of the Text S2 and S3 for details):

$$(M_R)_{ij} = p^R(g_i \to g_j)$$
$$(M_E)_{ij} = p^E(g_i \to g_j)$$

To determine the landscape crossing rate in a fluctuating environment, we compute the *mean first passage time,* the mean number of time steps before reaching the final genotype (76), t^{mfp}, of mutational trajectories going from genotype MK:acca = g_1 to genotype YQ:tggt = g_{64}. Because the paths stop at the latter genotype, for the computation we define it as an absorbing state in each environment, i.e.:

$$\left(M_{R/E}\right)_{64,j\neq 64} = p^{R/E}\left(\text{YQ:tggt} = g_{64} \rightarrow g_j \neq g_{64}\right) = 0$$

Next, from matrices M_R and M_E we extracted the submatrices Q_R and Q_E defined over all the 63 transient states (i.e. all genotypes $g_j \neq$ YQ:tggt, the absorbing genotype). Q_R and Q_E are thus 63×63 matrices. Q_R and Q_E are then used to compute the *fundamental matrix N(w,c)* of the evolution process (76) with as parameters *w*: the time fraction of environment *R,* and *c*: the number of successive mutation-fixation/extinction cycles occurring before any new possible change in environment is allowed (see Text S3). The environment is thus only allowed to change after a dwell-time of *c* mutation-fixation/extinction events (corresponding to *c* steps of the Markov process) and the new environment is picked with a probability *w* for environment *R* and a probability $1-w$ for environment *E*. The mean first passage time t^{mfp} and the corresponding rate k_c to evolve to YQ:tggt = g_{64} starting from MK:acca = g_1 are then obtained from the fundamental matrix $N(w,c)$ using equations S8 and S9 of the Text S4:

$$k_c(w,c) = 1/t^{mfp} = 1\Big/\sum_{transient\ states\ j} N_{1,j}(w,c)$$

Acknowledgements:

We thank D.M. Weinreich, J.A.G.M. de Visser, T. Paixão, J. Polechová, T. Friedlander and A.E. Mayo for reading and commenting on earlier versions of the manuscript, B. Houchmandzadeh, O. Rivoire and M. Hemery for discussions and suggestions on the Markov computation. Further, we kindly thank F.J. Poelwijk for sharing plasmid pCascade5 and pRD007 and Y. Yokobayashi for sharing plasmid pINV-110. We also thank the anonymous reviewers for their valuable remarks on the initial version of the manuscript. This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO). MGJ is co-funded by the VENI research programme, which is financed by the Netherlands Organisation for Scientific Research (NWO).

Statement of authorship:

MGJdV and SJT, conception and design of experiments. MGJdV and VS performed experiments. MGJdV, AD and SJT analysed the data. AD formalised and analysed the mathematical models. MGJdV, AD and SJT wrote the manuscript.

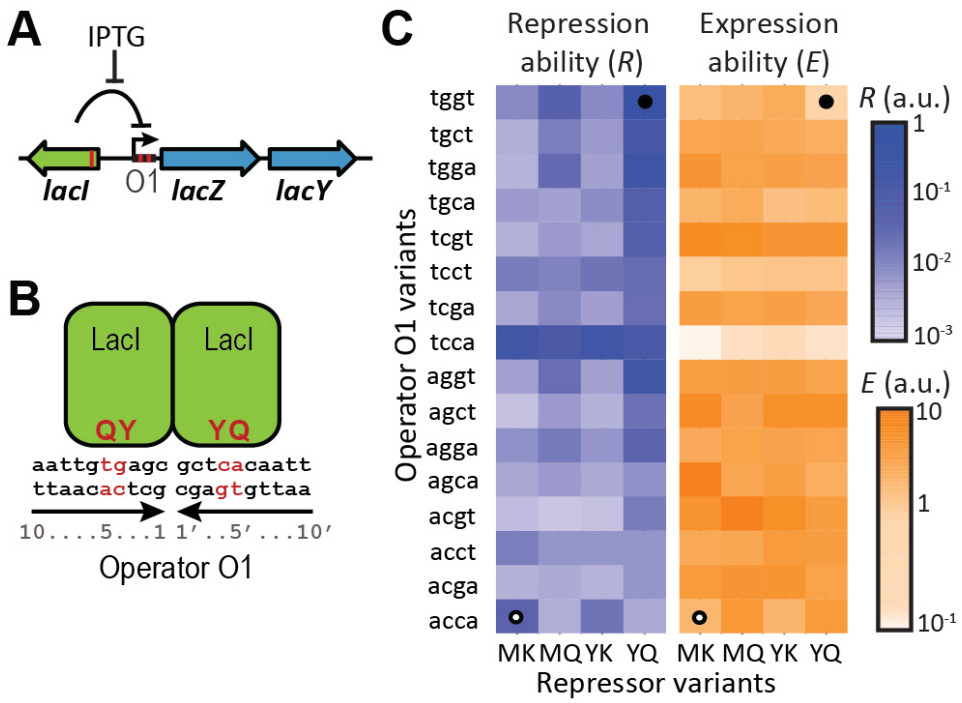

Figure 1 Repression and expression ability of *lac* repressor-operator mutants.

(A) Schematic representation of the *Escherichia coli lac* system. β-galactosidase (LacZ) and the *lac* permease (LacY) are co-regulated by the repressor LacI. Expression is induced by isopropyl-β-d-1-thiogalactopyranoside (IPTG). Red lines correspond to mutated positions. (B) The multimeric *lac* repressor in green bound to its operator DNA. Red: mutated positions responsible for specific repressor-operator binding. Y and Q are the mutated amino acid residues, on positions 17 and 18 in the DNA-binding helix of the *lac* repressor, 4 and 4' g-c and 5 and 5' t-a the mutated base-pairs in the operator DNA. We note the genotype represented here as YQ:tggt. (C) Characterization of the 64 *lac* repressor-operator variants. The starting and final sequences are indicated by open and filled circle respectively. Repression ability (*R*) is the inverse of the measured expression level in the absence of IPTG. Expression ability (*E*) is the measured expression level in the presence of IPTG (Methods).

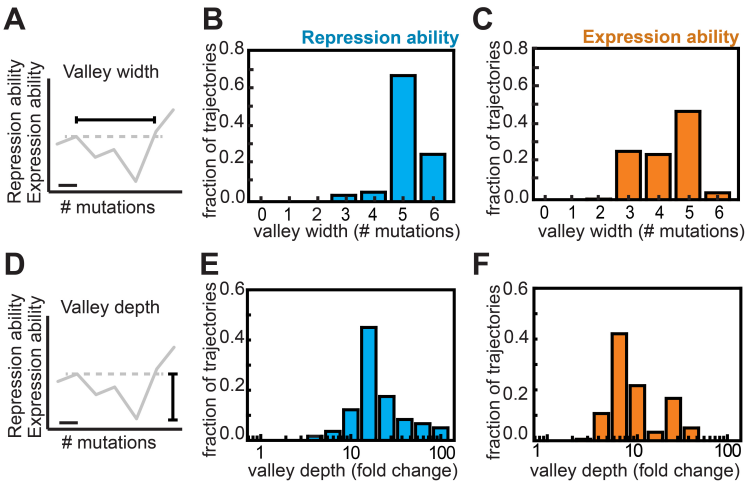

Figure 2 Genetic constraints in constant environments.

(A) The valley width is the number of mutations required to increase repression or expression ability above the previous suboptimum, in a given mutational trajectory. For each trajectory the widest valley is tabulated. Horizontal bar indicates one mutation. (B) Valley width in repression ability for all direct 720 mutational trajectories from MK:acca to YQ:tggt. (C) Valley width in expression ability. (D) The valley depth is the fold decrease in repression or expression ability within the valleys of a given trajectory. For each trajectory the deepest valley is tabulated. Note that the widest and the deepest valleys for a given trajectory do not necessarily coincide. (E) Valley depth in repression ability for all direct 720 mutational trajectories from MK:acca to YQ:tggt. (F) Valley depth in expression ability. Welch's t-tests were performed to determine the statistical significance of differences in measured repression and expression ability values (Methods).

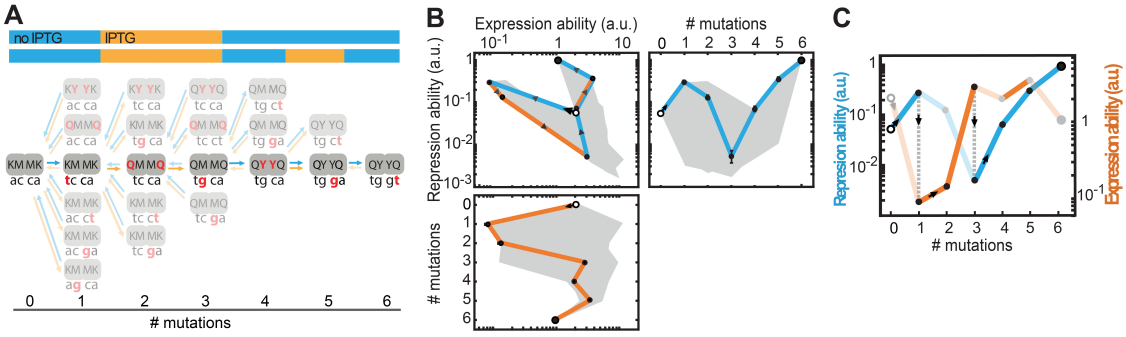

Figure 3 Escape from genetic constraint in fluctuating environments.

(A) Mutational trajectory accessible by continuous improvements in a changing environment, corresponding to panels B and C. Red indicates mutated position. Forward arrows indicate mutations conferring increases in repression or expression ability, backward arrows indicate decreasing or neutral steps. Top two color bars represent changing environments that confer continuous improvements. Without these changes the system would be trapped at MK:tcca and MQ:tgca, where no further improvements are possible in the current environment. (B) Expression and repression ability along the trajectory indicated in panel A. Data are represented as mean (N=3 or 4) ± SEM. The trajectory starts at MK:acca (open circle) and ends at YQ:tggt (large filled circle). This trajectory contains a valley in repression ability (top right panel, blue line) and in expression ability (bottom left panel, orange line). Top left panel: Blue lines indicate mutations that confer improvements in repression ability, orange lines indicate mutations that confer improvements in expression ability. The grey area indicates the envelope of all trajectories. (C) Schematic representation of repression and expression ability along the trajectory indicated in panel A in the variable environment. Semi-transparent lines indicate inaccessible mutations. Grey dotted lines indicate environmental changes allowing escape from sub-optima.

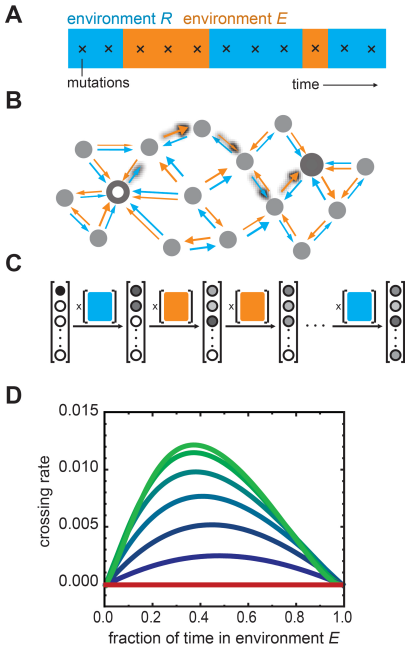

Figure 4 Landscape-crossing in stochastically alternating environments.

(A) Environmental fluctuations and occurrence of mutations (crosses) (Text S1). Environments *R* and *E* refer to the environment selecting respectively for repression and expression ability. (B) Schematic representation of genotype space. Large open and filled circles are start and end genotypes of mutational trajectories. Arrows indicate increasing repression ability (*R,* blue) or expression ability (*E,* orange); arrow thickness reflects magnitude and hence transition probability (Text S2). Shadowed arrows indicate one possible path of continuous improvement from the initial to the final genotypes. The structure of the space is schematic and does not reflect the actual system. (C) Schematic depiction of the Markov chain method for computing crossing rates. The probability vector lists all *N* genotypes, with the grey-scale indicating the probability of populating a genotype at a given indicated time. Initially only the beginning genotype is populated. The $N \times N$ environment-dependent transition probability matrices (colour squares) reflect the arrows in panel B: a matrix entry at position *i,j* indicates the transition probability from genotype *i* to genotype *j*. Each matrix multiplication yields a novel probability of genotype-occupancy after a mutation occurred in a given environment. This illustration is schematic: we use (an infinite time limit) analytical solution for this process considering a range of possible scenarios of environmental fluctuations. (D) Crossing rate k_c as a function of fraction of time spent in each environment, for different environmental fluctuation rates. The unit of time is the time between two mutations. Red line: environment dwell time << 1, meaning that the environment fluctuates much faster than the time between mutations. Top green line to bottom blue line: environment dwell time = 1, 2, 5, 10, 20, 50 (i.e. decreasing frequency of environmental fluctuation; Text S3-6). The crossing rate is the inverse of the mean number of mutations that are necessary to cross the landscape. The absolute maximum crossing rate is 0.17 (6^{-1}, corresponding to six mutations being fixed).

References

1. Weinreich DM, Delaney NF, Depristo MA, & Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111-114.
2. Poelwijk FJ, Kiviet DJ, Weinreich DM, & Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445(7126):383-386.
3. de Visser JA & Krug J (2014) Empirical fitness landscapes and the predictability of evolution. *Nature reviews. Genetics* 15(7):480-490.
4. Breen MS, Kemena C, Vlasov PK, Notredame C, & Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* 490(7421):535-538.
5. Kvitek DJ & Sherlock G (2011) Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. *PLoS genetics* 7(4):e1002056.
6. Salverda ML*, et al.* (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS genetics* 7(3):e1001321.
7. Poelwijk FJ, de Vos MGJ, & Tans SJ (2011) Tradeoffs and optimality in the evolution of gene regulation. *Cell* 146(3):462-470.
8. Meer MV, Kondrashov AS, Artzy-Randrup Y, & Kondrashov FA (2010) Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464(7286):279-282.
9. Miller SP, Lunzer M, & Dean AM (2006) Direct demonstration of an adaptive constraint. *Science* 314(5798):458-461.
10. Lind PA, Berg OG, & Andersson DI (2010) Mutational robustness of ribosomal protein genes. *Science* 330(6005):825-827.
11. Qu X*, et al.* (2011) Genetic epistasis between heparan sulfate and FGF-Ras signaling controls lens development. *Developmental biology* 355(1):12-20.
12. Via S (2002) The ecological genetics of speciation. *Am Nat* 159 Suppl 3:S1-7.
13. Schluter D & Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl 1:9955-9962.
14. Dettman JR, Sirjusingh C, Kohn LM, & Anderson JB (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature* 447(7144):585-588.
15. Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, & Tans SJ (2011) Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J Theor Biol* 272(1):141-144.
16. Lozovsky ER*, et al.* (2009) Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America* 106(29):12025-12030.
17. Macken CA & Perelson AS (1989) Protein evolution on rugged landscapes. *Proceedings of the National Academy of Sciences of the United States of America* 86(16):6191-6195.
18. Burton OJ & Travis JMJ (2008) Landscape structure and boundary effects determine the fate of mutations occurring during range expansions. *Heredity* 101:329-340.
19. Bitbol AF & Schwab DJ (2014) Quantifying the role of population subdivision in evolution on rugged fitness landscapes. *PLoS computational biology* 10(8):e1003778.
20. Weinreich DM & Chao L (2005) Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution Int J Org Evolution* 59(6):1175-1182.
21. de Visser JA, Park SC, & Krug J (2009) Exploring the effect of sex on empirical fitness landscapes. *Am Nat.* 174:S15-S30.

22. Watson RA, Weinreich DM, & Wakeley J (2010) Genome structure and the benefit of sex. *Evolution* 65(2):523-536.
23. Watson RA, Weinreich DM, & Wakeley J (2006) Effects of intra-gene fitness interactions on the benefit of sexual recombination. *Biochem Soc Trans* 34(Pt 4):560-561.
24. Kimura M (1983) *The Neutral Theory of Molecular Evolution.* (Cambridge University Press, Cambridge).
25. Lynch M (2007) The evolution of genetic networks by non-adaptive processes. *Nature reviews. Genetics* 8(10):803-813.
26. Weissman DW, Desai MM, Fisher DS, & Feldman MW (2009) The Rate at Which Asexual Populations Cross Fitness Valleys. *Theoretical Population Biology* 75:286-300
27. Eldar A*, et al.* (2009) Partial penetrance facilitates developmental evolution in bacteria. *Nature* 460(7254):510-514.
28. Crispo E (2007) The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution* 61(11):2469-2479.
29. Mills R & Watson RA (2006) On Crossing Fitness Valleys with the Baldwin Effect. *Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*:493-499.
30. Remold SK & Lenski RE (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in Escherichia coli. *Nature genetics* 36(4):423-426.
31. Bataillon T, Zhang T, & Kassen R (2011) Cost of Adaptation and Fitness Effects of Beneficial Mutations in Pseudomonas fluorescens. *Genetics*.
32. Hall AR, Iles JC, & MacLean RC (2011) The fitness cost of rifampicin resistance in *Pseudomonas aeruginosa* depends on demand for RNA polymerase. *Genetics* 187(3):817-822.
33. Stiffler MA, Hekstra DR, & Ranganathan R (2015) Evolvability as a Function of Purifying Selection in TEM-1 beta-Lactamase. *Cell* 160(5):882-892.
34. de Vos MG, Poelwijk FJ, Battich N, Ndika JD, & Tans SJ (2013) Environmental dependence of genetic constraint. *PLoS genetics* 9(6):e1003580.
35. Flynn KM, Cooper TF, Moore FB, & Cooper VS (2013) The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS genetics* 9(4):e1003426.
36. Lindsey HA, Gallie J, Taylor S, & Kerr B (2013) Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* 494(7438):463-467.
37. Kashtan N, Noor E, & Alon U (2007) Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104(34):13711-13716.
38. Tan L, Serene S, Chao HX, & Gore J (2011) Hidden randomness between fitness landscapes limits reverse evolution. *Phys Rev Lett* 106(19):198102.
39. Mustonen V & Lässig M (2008) Molecular Evolution under Fitness Fluctuations. *Physical Review Letters* 100:108101-108101- 108101-108104.
40. Mustonen V & Lassig M (2009) From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends in genetics : TIG* 25(3):111-119.
41. Dekel E & Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436(7050):588-592.
42. Eames M & Kortemme T (2012) Cost-benefit tradeoffs in engineered *lac* operons. *Science* 336(6083):911-915.
43. Nguyen CC & Saier MH, Jr. (1995) Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS Lett* 377(2):98-102.
44. Weickert MJ & Adhya S (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J Biol Chem* 267(22):15869-15874.

45. Sartorius J, Lehming N, Kisters B, von Wilcken-Bergmann B, & Muller-Hill B (1989) *lac* repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. *Embo J* 8(4):1265-1270.
46. Lehming N, Sartorius J, Kisters-Woike B, von Wilcken-Bergmann B, & Muller-Hill B (1990) Mutant *lac* repressors with new specificities hint at rules for protein--DNA recognition. *Embo J* 9(3):615-621.
47. Lewis M (2013) Allostery and the *lac* Operon. *Journal of molecular biology* 425(13):2309-2316.
48. Draghi JA, Parsons TL, Wagner GP, & Plotkin JB (2010) Mutational robustness can facilitate adaptation. *Nature* 463(7279):353-355.
49. Gillespie JH (1983) Some properties of finite populations experiencing strong selection and weak mutation. *Am Nat* 121:691-708.
50. McCandlish DM (2013) On the findability of genotypes. *Evolution* 67(9):2592-2603.
51. Perfeito L, Ghozzi S, Berg J, Schnetz K, & Lassig M (2011) Nonlinear fitness landscape of a molecular pathway. *PLoS genetics* 7(7):e1002160.
52. Burch CL & Chao L (1999) Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics* 151(3):921-927.
53. Nahum JR*, et al.* (2015) A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*.
54. Wagner GP & Lynch VJ (2010) Evolutionary novelties. *Current biology : CB* 20(2):R48-52.
55. Wu CI & Ting CT (2004) Genes and speciation. *Nature reviews. Genetics* 5(2):114-122.
56. Stearns SC (1989) Trade-offs in life-history evolution. *Functional Ecology* 3(3):259-268.
57. Levins R (1968) *Evolution in Changing Environments* (Princeton University Press, Princeton).
58. Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365(1537):87-97.
59. Kim S, Lieberman TD, & Kishony R (2014) Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. *Proceedings of the National Academy of Sciences of the United States of America* 111(40):14494-14499.
60. Schenk MF*, et al.* (2015) Role of pleiotropy during adaptation of TEM-1 beta-lactamase to two novel antibiotics. *Evolutionary applications* 8(3):248-260.
61. Fuentes-Hernandez A*, et al.* (2015) Using a sequential regimen to eliminate bacteria at sublethal antibiotic dosages. *PLoS biology* 13(4):e1002104.
62. Haldimann A, Daniels LL, & Wanner BL (1998) Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *Journal of bacteriology* 180(5):1277-1286.
63. Yokobayashi Y, Weiss R, & Arnold FH (2002) Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences of the United States of America* 99(26):16587-16591.
64. Amann E, Ochs B, & Abel KJ (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* 69(2):301-315.
65. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, & Muller-Hill B (1994) Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. *Embo J* 13(14):3348-3355.
66. Oehler S (2009) Feedback regulation of Lac repressor expression in *Escherichia coli*. *Journal of bacteriology* 191(16):5301-5303.
67. Marbach A & Bettenbrock K (2012) *lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of biotechnology* 157(1):82-88.
68. Roderick SL (2005) The *lac* operon galactoside acetyltransferase. *C. R. Biologies* 328:568-575.
69. Simons A, Tils D, von Wilcken-Bergmann B, & Muller-Hill B (1984) Possible ideal lac operator: Escherichia coli lac operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proceedings of the National Academy of Sciences of the United States of America* 81(6):1624-1628.

70. Grinstead C & Snell L (2006) Grinstead and Snell's Introduction to Probability. *American Mathematical Society*.
71. Gillespie JH (1983) A simple stochastic gene substitution model. *Theor Popul Biol* 23(2):202-215.
72. Houchmandzadeh B & Vallade M (2010) Alternative to the diffusion equation in population genetics. *Physical review. E, Statistical, nonlinear, and soft matter physics* 82(5 Pt 1):051913.
73. DePristo MA, Hartl DL, & Weinreich DM (2007) Mutational reversions during adaptive protein evolution. *Molecular biology and evolution* 24(8):1608-1610.