


ARTICLE

DOI: 10.1038/s42003-018-0160-1

OPEN

Fitness variation in isogenic populations leads to a novel evolutionary mechanism for crossing fitness valleys

Debra Van Egeren ^{1,2}, Thomas Madsen^{1,3} & Franziska Michor ^{1,3,4,5,6,7}

Individuals in a population often have different fitnesses even when they have identical genotypes, but the effect of this variation on the evolution of a population through complicated fitness landscapes is unknown. Here, we investigate how populations with non-genetic fitness variation cross fitness valleys, common barriers to adaptation in rugged fitness landscapes in which a population must pass through a deleterious intermediate to arrive at a final advantageous stage. We develop a stochastic computational model describing the dynamics of an asexually reproducing population crossing a fitness valley, in which individuals of the same evolutionary stage can have variable fitnesses. We find that fitness variation that persists over multiple generations increases the rate of valley crossing through a novel evolutionary mechanism different from previously characterized mechanisms such as stochastic tunneling. By reducing the strength of selection against deleterious intermediates, persistent fitness variation allows for faster adaptation through rugged fitness landscapes.

¹Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ²Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. ³Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02215, USA. ⁴Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. ⁵The Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ⁶The Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA. ⁷The Ludwig Center at Harvard, Boston, MA 02115, USA. Correspondence and requests for materials should be addressed to F.M. (email: michor@jimmy.harvard.edu)

Variation in reproductive fitness among individuals of a population is common. In some cases, this variation is due to genetic heterogeneity, where multiple segregating mutations with different fitness effects exist simultaneously in the population^{1,2}. However, recent work quantifying interdivision times in isogenic mammalian cell populations demonstrated that fitness variation can exist even among individuals with the same genotype^{3,4} (Fig. 1a, b). This non-genetic fitness variation may originate from several different intrinsic and extrinsic sources, each of which can have different persistence timescales. For example, gene expression heterogeneity has the potential to modify a cell's fitness⁵ and is a source of non-genetic resistance in cancer⁶. These transcriptional fluctuations have been shown to persist over 3–4 human cell divisions⁷. Epigenetic modifications such as DNA methylation persist over longer timescales (>20 cell divisions)⁸ and may represent an additional source of stable fitness variation in genetically identical individuals. Here, we investigated how this non-genetic fitness variation affects evolution through complex fitness landscapes, and how these effects depend on the magnitude and persistence length of fitness variation.

Background fitness variation has previously been shown to affect evolutionary trajectories by reducing the strength of selection for or against new mutations. This phenomenon is known as clonal interference or the Hill–Robertson effect^{9,10} in the context of genetic fitness variation. When new mutations arise, they are linked to the genetic background of the individual in which the mutation occurred. If a population includes many different genotypes, this linkage leads to greater variability in the relative fitness of new mutants, reducing the average magnitude of the fitness effect of the mutation and leading to weaker selection^{11–13}. In sexually reproducing populations, genetic recombination breaks the linkage between the mutation of

interest and other loci in the original genotype, removing any fitness effect conferred by that genotype in subsequent generations¹⁴. While the Hill–Robertson effect has been extensively investigated in the context of a single locus or multiple independent loci, the effect of fitness variation on populations in rugged fitness landscapes with sign epistasis is less well understood. Here, we investigate how stable fitness variation affects the evolution of populations crossing fitness valleys, which are frequently found in complex fitness landscapes.

Fitness valleys are barriers to adaptation that exist when a population of individuals can acquire an advantageous trait only by passing through an intermediate stage of lower fitness (Fig. 1c). During tumorigenesis, for example, fitness valleys are represented by situations in which cells that lost one functional copy of a tumor suppressor gene are less fit than the original population¹⁵. Other examples arise during the development of antibiotic resistance in bacteria¹⁶, and immune system escape in influenza¹⁷ and HIV¹⁸, and affect the speed and trajectory of evolutionary adaptation in these populations. In particular, selection against the deleterious intermediate stage decreases the rate at which individuals of the final, advantageous evolutionary stage arise, rendering adaptation to the fitness peak difficult. Previous work¹⁹ characterizing the dynamics of valley crossing identified two mechanisms by which populations can overcome this adaptation barrier (Fig. 1d). First, in smaller populations with a weakly deleterious intermediate, the intermediate stage may reach 100% frequency via drift, after which final-stage individuals emerge and sweep throughout the population. This evolutionary mechanism is referred to as sequential fixation. However, increasing the population size or the valley depth decreases the chance of the population evolving along this route. In these cases, valley crossing often occurs through a second mechanism known as stochastic tunneling, in which final-stage individuals emerge

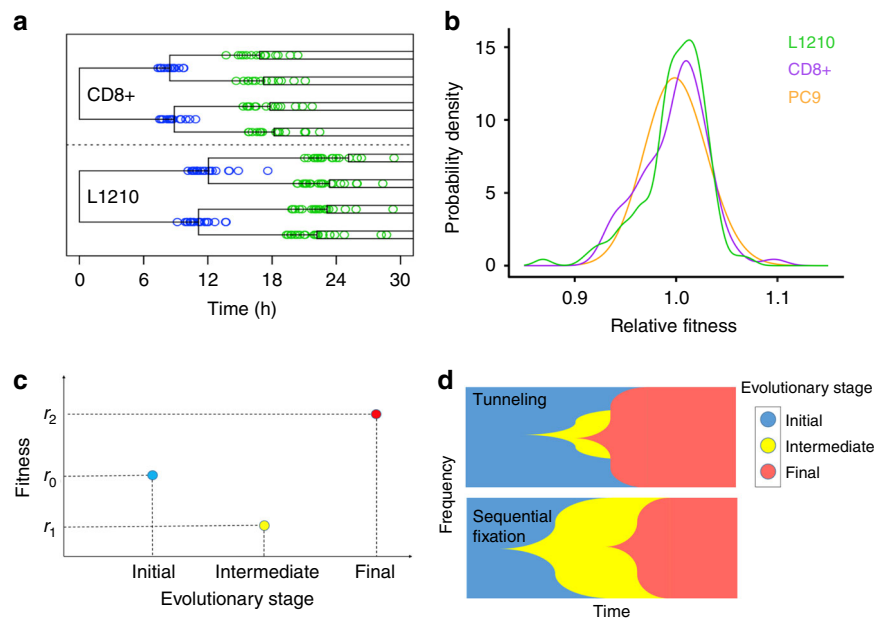


Fig. 1 Variable fitness values in isogenic populations and mechanisms of adaptation to rugged fitness landscapes. **a** Single cell interdivision times in lineages of primary murine CD8⁺ T cells and a murine lymphocytic leukemia cell line (L1210). Division times of cells in each generation (green and blue circles) are variable, as shown by their distribution across the time axis. **b** Fitness distributions estimated from single cell PC9 (a human lung cancer cell line), CD8⁺, and L1210 interdivision time data. Fitness is estimated as the reciprocal of the interdivision time. All distributions are scaled to have mean 1 and equal variance. **c** Illustration of a fitness valley. Individuals of the initial evolutionary stage have fitness r_0 , individuals of the intermediate stage have a lower fitness r_1 , and individuals of the final stage have a higher fitness r_2 . **d** Illustration of mechanisms of fitness valley crossing. If stochastic tunneling occurs, individuals of the final evolutionary stage emerge before the individuals of the intermediate stage take over the population (top). Without tunneling, the disadvantageous intermediate type fixes before the final advantageous trait emerges and fixes in the population (bottom)

and reach fixation in the population before the deleterious intermediate does^{20–22}. The rates of both of these valley crossing mechanisms have been derived for asexually reproducing populations without background genetic variation²³.

The dynamics of fitness valley crossing has previously been shown in models of viral dynamics to be affected by genetic clonal interference and recombination. Neher and Shraiman derived the rate of fixation of new mutations and the rate of stochastic tunneling for a mathematical model of HIV replication and recombination in which mutations are common and clonal interference is frequent²⁴. They showed that the rate of fixation of the deleterious intermediate and the rate of stochastic tunneling are higher in populations with more clonal interference, i.e., low recombination rates and high background fitness variation, leading to an increased rate of fitness valley crossing. We hypothesized that a similar increase in valley crossing rates also results from non-genetic variation in genetically identical individuals. While prior theoretical work suggests that a similar reduction in selection strength results from noise in gene expression levels²⁵, the effect of non-genetic background fitness on evolution in more complex fitness landscapes has not yet been systematically explored.

Here, we design a computational model in which small, frequent stochastic fitness alterations generate population-level fitness variation, recapitulating many of the features observed in isogenic cell populations. Using this model, we find that non-genetic fitness variation that persists across multiple generations increases the rate of valley crossing. This increase in the rate of valley crossing occurs via a new mechanism in which the emergence of individuals of the intermediate stage with high-fitness backgrounds increases the rate of intermediate-stage fixation. Additionally, we show that the adaptation rate is increased as the magnitude of the background non-genetic fitness variation increases or as the persistence timescale of the fitness alterations lengthens. Therefore, by reducing the efficacy of selection against weakly deleterious intermediate traits, stable background fitness variation increases the rate of adaptation in rugged fitness landscapes and leads to a novel evolutionary mechanism of crossing fitness valleys.

Results

Fitness alterations lead to steady-state fitness variation. To investigate the role of background fitness variation on the dynamics of valley crossing, we considered a modified Moran process²⁶ in which asexually reproducing individuals acquire (epi)genetic alterations that change their fitness during every reproductive event (Fig. 2a, Methods). During its i th reproductive event, an individual acquires a new multiplicative fitness effect $m_i \sim F(1, V)$ drawn from a distribution of fitness effects centered around neutrality ($m_i = 1$) with variance V . Each stochastic fitness alteration persists through exactly τ cell divisions, including the division in which the alteration was acquired. If fitness effects are permanent ($\tau = \infty$), the mean fitness of the population increases without bound due to selection; otherwise, the mean population fitness fluctuates around a steady-state level (Fig. 2b). The width of the steady-state relative fitness distribution increases with the variance of the fitness effect distribution, V (Supplementary Figure 1) and with the persistence timescale τ of fitness alterations (Fig. 2c). Additionally, as fitness alterations become more persistent, i.e., as τ increases, the shape of the steady-state fitness distribution becomes less sensitive to the shape of the fitness effect distribution, F (Supplementary Figure 2a–c). We found that four different fitness effect distributions F (log-normal, gamma, exponential, and centered Bernoulli) resulted in steady-state fitness distributions that are similar to each other (Methods, Supplementary Figure 2d). We used a log-normal distribution for

F for the remainder of our investigations as a representative example.

Using this model, we found that the fitness distributions in simulated populations with non-genetic fitness variation were similar to fitness distributions measured in isogenic single cell experiments (Fig. 2d). Using single-cell interdivision time data, we estimated the fitness distribution of three types of cells: a human non-small cell lung cancer cell line (PC9)²⁷, primary murine CD8+ T cells, and a murine lymphocytic leukemia cell line (L1210)⁴ (Methods). We found that these experimentally derived distributions were not significantly different from our simulated distributions with fitness effects that persist over 10 generations ($p = 0.99$ for both CD8+ and L1210 fitness distributions; Methods). Results from our variable fitness model are, therefore, consistent with the experimentally observed fitness distributions. However, multiple different sets of simulation parameter values (τ , F , and V) result in similar population fitness distributions. Therefore, we were unable to infer unique combinations of τ , F , and V that would best match the experimental data, but some of these parameter values are more likely to be biologically relevant than others, such as τ on the order of 3–20 generations^{7,8}.

Stable fitness variation promotes valley crossing. To investigate how populations with non-genetic fitness variation cross rugged fitness landscapes, we specified a two-step evolutionary model defining a fitness valley (Fig. 3a, Methods). During each reproduction event, individuals draw new random fitness effects as described above, but also may mutate to the next evolutionary stage. The population begins in the initial evolutionary stage S_0 with average fitness r_0 . Here, we specify $r_0 = 1$ for simplicity. Individuals of stage S_0 can acquire a mutation to transition to the intermediate stage S_1 with probability u_1 during each reproductive event. This transition to the intermediate stage is associated with a multiplicative fitness cost $r_1/r_0 < 1$. Thus an S_0 individual with fitness w after the stochastic fitness alteration process gives rise to a mutated S_1 daughter with fitness $w \times (r_1/r_0)$. Intermediate-stage individuals can then mutate to the final, advantageous evolutionary stage with probability u_2 per reproductive event. The fitness of the final-stage relative to that of the initial S_0 individuals is given by $r_2 > r_0$, and the fitness benefit conferred by the transition from S_1 to S_2 is therefore $r_2/r_1 > 1$. We implemented this model as a stochastic computer simulation to determine the rate of crossing fitness valleys and of stochastic tunneling in populations with intra-stage fitness variability.

We found that populations with fitness variation that is stable over multiple generations cross fitness valleys more quickly than populations with no persistent fitness variation (Fig. 3b). In our model, we assume that there is no back mutation since many mutations that affect fitness are loss-of-function mutations which have a very low reversal rate. Therefore, the final evolutionary stage will eventually reach fixation in all populations. However, the state of populations at intermediate timescales differ depending on their non-genetic fitness variation properties. Here, we focused on the state of populations after 3650 generations, corresponding to timescales relevant for tumorigenesis²⁸, and found that this increased rate of adaptation strongly depends on the fitness effect persistence timescale τ (Fig. 3c). Using the parameter values shown in Table 1, if stochastic fitness alterations only persist for approximately 1–10 generations, intra-stage fitness variation does not result in valley crossing rates that are different from those observed in the zero-variance model with no intra-stage fitness variation. Additionally, populations with stable fitness variation that acquire the final, advantageous trait are more likely to cross the fitness valley without tunneling (Fig. 3d).

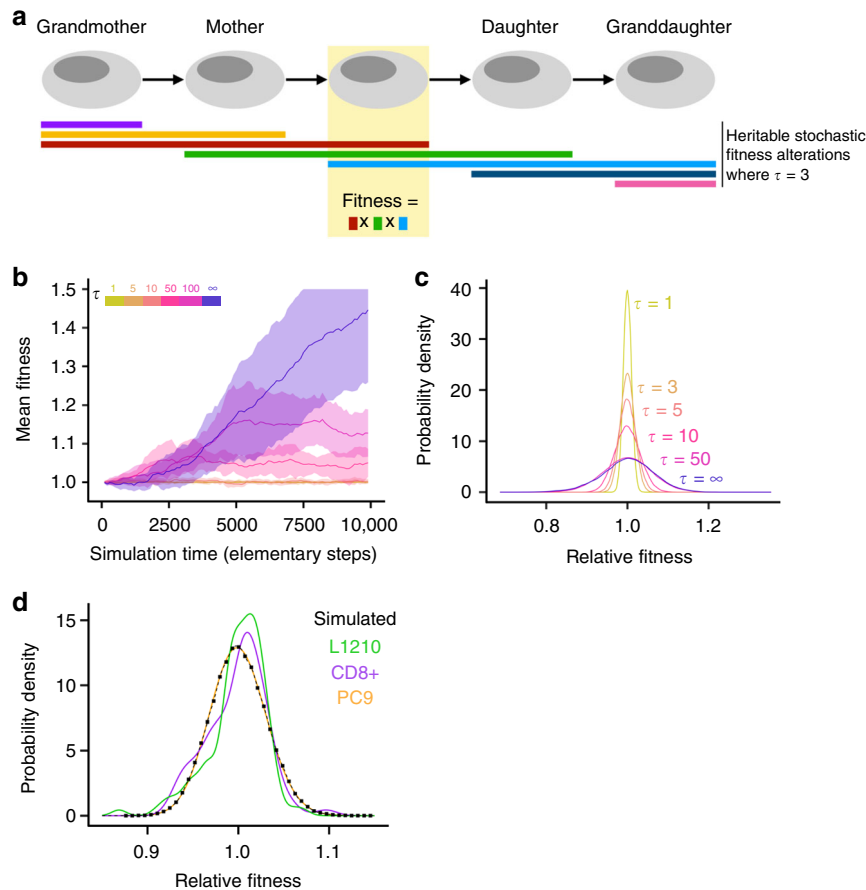


Fig. 2 An evolutionary model incorporating non-genetic fitness changes can generate steady-state fitness variation similar to that seen in isogenic cell populations. **a** Description of the mathematical model. During a cell's i th division, it acquires an epigenetic alteration (colored rectangles) that changes its fitness by effect size m_i . This alteration is carried through the cell's next $\tau - 1$ divisions. **b** Mean fitness of the population reaches a steady-state level for finite fitness effect lifetimes τ but diverges when non-genetic fitness changes persist indefinitely ($\tau = \infty$). Shaded regions are one simulated standard deviation above and below the mean fitness at each point in simulation time ($n = 10$ independent simulations). **c** Effect of τ on the simulated steady-state fitness distribution of individuals of a single evolutionary stage. Increasing τ increases the width of the fitness distribution. Distributions are estimated from 1000 independent simulations of $N = 100$ individuals after 3600 generations. **d** Comparison of the simulated steady-state fitness distribution of a population with experimentally derived fitness distributions. The black simulated distribution with fitness variation that persists over 10 generations ($\tau = 10$) is similar to properly scaled and shifted versions of the empirically estimated fitness distributions (orange: PC9/log-normal, purple: CD8+, green: L1210). The differences between the simulated distribution and the experimental CD8+ distribution ($p = 0.99$, $Z = -0.0066$) and between the simulated distribution and the experimental L1210 distribution ($p = 0.99$, $Z = 0.0089$) are not statistically significant. Simulated distributions are combined data from 1000 separate simulations (100,000 total individuals) after 3600 generations. Simulation parameter values in this figure not otherwise specified are the standard values given in Table 1

These observations suggest that individuals with the disadvantageous intermediate trait are more likely to fix when there is persistent fitness variation in the population, which represents a new, tunneling-independent mechanism of valley crossing.

Our results are robust to the shape of the fitness effect distribution (Supplementary Figure 3) and the mutation rates between the evolutionary stages (Supplementary Figure 4). The mean relative fitness advantage of the final stage affects valley crossing trajectories for lower values of the fitness advantage r_2 , for which drift plays a more important role in the fixation of the final evolutionary stage (Supplementary Figure 5). The effect of persistent variation is most pronounced in intermediately sized populations (Supplementary Figure 6). In large populations (thousands of cells or more), stochastic tunneling is inevitably the fastest trajectory for crossing the fitness valley. In smaller populations, drift dominates and reduces the effects of selection, leading to frequent fixation of the intermediate stage and reducing the incidence of tunneling. Finally, this increase in the valley crossing rate is observed in populations in which

non-genetic fitness effects do not have fixed lifetimes, but rather have a fixed probability of reversion at every reproductive event, leading to geometrically distributed effect lifetimes (Supplementary Figure 7, Methods). In this model, even fitness effects that are expected to persist only for a few generations (>3 generations) have an impact on valley crossing dynamics, which is in contrast to our results from the deterministic lifetime model (Fig. 3c). This difference between the deterministic and geometrically distributed non-genetic effect lifetime models is likely due to very long-lived fitness effects from the heavy tail of the geometric lifetime distribution that persist beyond the mean lifetime τ .

Fitness effect parameters affect valley crossing dynamics. We hypothesized that an association of the disadvantageous trait with a high-fitness background could explain the increase in the fixation probability of the intermediate stage observed in populations with fitness variation. If the disadvantageous trait arises in a particularly fit individual in the initial population that has

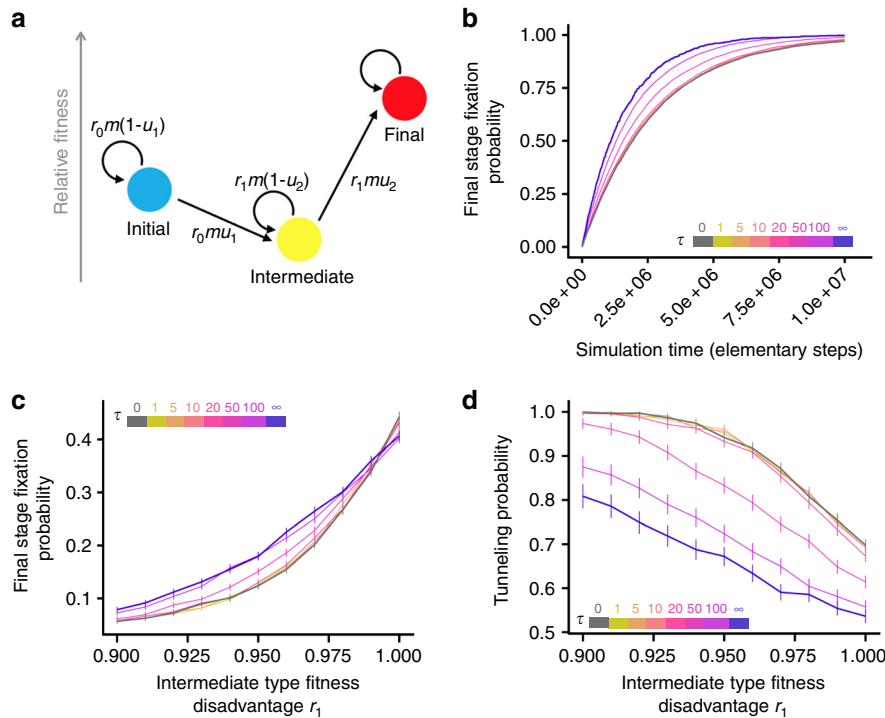


Fig. 3 Persistent variation in fitness increases the rate of valley crossing by increasing the rate of intermediate stage fixation. **a** The definition of a fitness valley in the evolutionary model. Transition rates shown in the diagram are defined for an individual with a total stochastic fitness effect m . **b** Rate of valley crossing for populations with fitness alterations with different lifetimes τ . Less time is required for the final evolutionary stage to fix in the population as τ increases. Here, 10,000 simulations per model were run. **c** Simulated probability of crossing the fitness valley and reaching fixation of the final, advantageous evolutionary stage within 3650 generations across different valley depths r_1 . A cell generation is N (the population size, here 100 individuals) elementary steps; one elementary step consists of one cell dying and a new cell being born to replace it. 10,000 simulations per data point. Vertical bars are 95% confidence intervals. Line colors represent values of τ , as in **b**. **d** Tunneling probability across different valley depths r_1 . The tunneling probability is calculated as the conditional probability that tunneling occurs, given that the final advantageous stage fixes during the time of the simulation. 10,000 simulations per data point; vertical bars are 95% confidence intervals. Simulation parameter values in this figure not otherwise specified are the standard values given in Table 1

Table 1 Simulation parameter values

Parameter	Value	Description
N	100	Number of individuals
F	Log-normal	Fitness effect distribution
V	1×10^{-4}	Variance of F
u_1	1×10^{-4}	Initial to intermediate-stage mutation rate
u_2	2×10^{-4}	Intermediate to final-stage mutation rate
r_0	1	Initial stage starting mean fitness
r_1	0.95	Intermediate stage fitness disadvantage
r_2	5	Final-stage fitness advantage
T_{max}	$3650(N)$	Truncation time (in elementary steps)

The definitions and default values chosen for each parameter of our model. All simulations were performed with these values unless otherwise specified

accumulated stable beneficial fitness alterations, it will pass on that advantage to its descendants, increasing the chance of fixation of that lineage. If associations arising by chance between beneficial fitness backgrounds and the disadvantageous trait explain our results, there should be two important characteristics that affect the incidence of this novel valley crossing mechanism in a population with fitness variation: First, the frequency of emergence of intermediate-stage individuals with high-fitness backgrounds; and second, the efficacy with which these individuals pass their fitness advantage on to their descendants.

The ability of a population to generate new intermediate-stage individuals with relatively high fitnesses depends on the variance V of the distribution of fitness effects. Larger values of V lead to greater variability in the fitness of new intermediate-stage mutants, in part by increasing the steady-state population variance. Increasing V will therefore increase the chance of generating particularly high-fitness intermediate individuals, ultimately leading to higher rates of valley crossing without stochastic tunneling. This prediction is validated by our simulation data (Fig. 4a, b, Supplementary Figure 8). We also found that populations with larger values of V are able to cross deeper fitness valleys, again because they are able to produce a small number of intermediates with relatively high fitnesses despite large fitness disadvantages due to the deleterious mutation.

The fitness effect persistence lifetime τ influences the fixation probability of intermediate-stage individuals by modulating the effect of a beneficial fitness background of a new intermediate individual. Lower values of τ (less persistent non-genetic fitness alterations) imply that any background fitness gains present in a particular intermediate-stage individual will not be passed down to many of its descendants and will not persist over the long timescales needed to substantially affect its overall fixation probability. Indeed, we found that for lower values of τ , intermediate individuals that initially have high relative fitness are about as likely to generate a successful lineage as individuals with lower initial fitness (Fig. 4c). However, as τ increases, intermediate-stage individuals with higher relative fitnesses

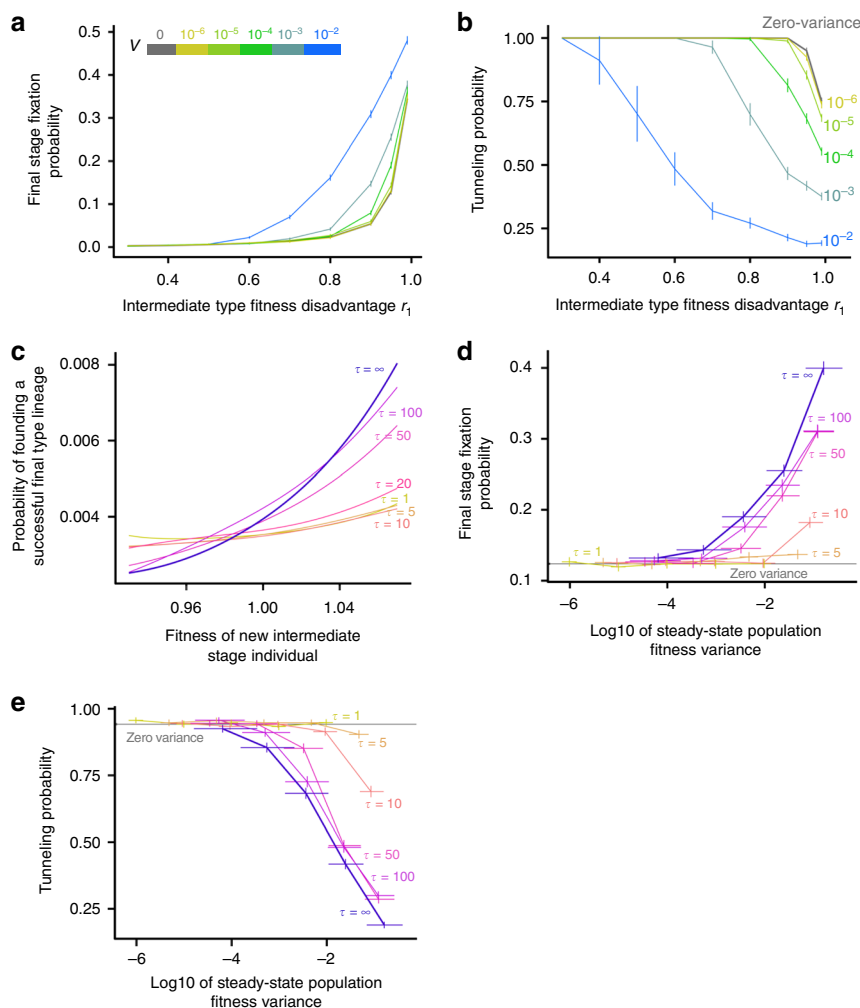


Fig. 4 Populations with epigenetic fitness alterations that are larger or persist over more generations can cross deeper fitness valleys. **a** Probability that the final advantageous stage will fix after 3650 generations with different variances for the fitness effect distributions V . Different colors represent different values of V , with the same color mapping marked for the traces in **b**. **b** Tunneling probability increases for larger values of V (values marked to the right of each line) for equivalent values of the valley depth r_1 . 10,000 simulations per data point; vertical bars are 95% confidence intervals. **c** Smoothed probability that a new intermediate-stage individual gives rise to a final-stage individual that eventually fixes in the population. When fitness changes persist over more generations (higher values of τ), intermediate individuals that arise with a higher relative fitness are more likely to establish a lineage that will eventually fix. Details on how this conditional probability was estimated are given in Methods. **d** Populations with more stable fitness effects are more likely to cross fitness valleys, even if they have the same steady-state fitness variance. **e** The decrease in the tunneling probability observed for populations with higher τ is not dependent on the increase in steady-state distribution width associated with more stable fitness effects. Error bars are 95% confidence intervals. 10,000 simulations per point. Steady-state fitness distribution variances were estimated from 1000 independently simulated distributions. Simulation parameter values in this figure not otherwise specified are the standard values given in Table 1

become more likely to establish a successful final-stage lineage. The beneficial fitness alterations present in these intermediate individuals are able to persist through several reproductive events and ultimately have a greater impact on the fate of the new lineage. The stability of the fitness effects across multiple generations remains an important determinant of valley crossing trajectories even after controlling for the increase in the steady-state fitness distribution width at higher values of τ (Fig. 4d, e).

Discussion

Here, we characterized a mathematical model describing the rate and mechanisms of adaptation to rugged fitness landscapes of populations with inter-individual variation in fitness. Using this model, we showed that populations with fitness variation that persists across multiple generations can cross fitness valleys more quickly than populations without such variation; the evolutionary

mechanism by which these populations adapt to such landscapes allows them to fix the deleterious intermediate stage by hitchhiking on stochastically accumulated high-fitness backgrounds. The magnitude of this effect strongly depends on the number of generations over which fitness variation persists, the valley depth, and the variance of the fitness effect distribution. Our analyses revealed that populations with frequent fitness alterations mediated by epigenetic mechanisms that are stable over multiple generations are more likely to cross fitness valleys without stochastic tunneling.

The dependence of this effect on the fitness effect lifetime is supported by previous work on similar evolutionary phenomena. Prior characterizations of the Hill–Robertson effect emphasize that recombination reduces the impact of a variable fitness background on the fate of a linked mutation^{9,10}. In those contributions, the recombination rate effects are related to the timescale of association of the fitness background with the genetic

mutations that specify the fitness valley—more recombination implies that the fitness background has less of an impact on the fate of a mutation over the time period during which it is segregating in the population. Therefore, if the linkage between fitness background and the trait in question is weak, theoretical descriptions of Hill–Robertson interference predict that the effect of fitness variation on selection is minimal. Similar observations were made in our model when τ was small. Note that the effect of the fitness effect lifetime τ on the adaptation rate is modulated by other parameters in our model such as the amount of standing fitness variation in the population. Therefore, there is no constant critical threshold of τ for all populations below which stochastic fitness alterations have a negligible impact on evolution.

The depth of fitness valleys that can be crossed with our observed evolutionary mechanism increases with enhanced fitness variation. Many deleterious mutations have low selection coefficients (corresponding to valley depths $r_1/r_0 > 0.9$) as determined experimentally^{1,29}, corresponding to the parameter regime in which we observe this effect. Therefore, our proposed mechanism for crossing fitness valleys may affect all evolutionary trajectories through these deleterious states even at minimal fitness distribution variances ($V \sim 10^{-4}$ – 10^{-5}) by increasing the rate of fixation of deleterious mutations.

The population size N is particularly important for determining drift-selection dynamics relevant to valley crossing situations. Previous work has demonstrated that intermediately sized populations are governed by ‘one-hit’ dynamics when crossing fitness valleys¹⁹. By accelerating the rate-limiting step of generating a substantially sized population of intermediate-stage individuals, persistent fitness variation speeds up adaptation in this range of population sizes. Such situations are of special interest in carcinogenesis; for instance, colonic crypts are populated by a small number of stem cells³⁰. In such situations, increasing the amount of fitness variation (for example by acquiring a mutator phenotype) may increase the rate of fixing mutations with mildly deleterious fitness effects that can nevertheless lead to further adaptation.

While we demonstrated that stable fitness variation increases the rate of adaptation of a population in a specific fitness landscape topology (fitness valleys), this type of heterogeneity also has important consequences that affect evolution through other fitness landscapes. One immediate consequence of the observed increased fixation probability of weakly deleterious mutations is an increased deleterious mutational load. These deleterious mutations may potentiate further adaptation by acting as intermediate stages in fitness valleys but may also lower the overall population fitness. Furthermore, just as background fitness variation reduces the strength of selection against deleterious mutations, it also reduces the strength of selection for weakly beneficial mutations, possibly hindering adaptation. These effects may influence the way heterogeneous populations move through fitness landscapes. Therefore, further theoretical and experimental studies on more general fitness landscapes are warranted and will lead to additional insight into the consequences of non-genetic sources of inter-individual fitness variation.

Methods

Computational evolutionary model with fitness variation. Our model is a variation of the standard Moran model of evolutionary dynamics that describes fitness valley crossing with different amounts of fitness variation that persist over one or more generations. In the Moran model, S_0 , S_1 , and S_2 denote the initial, intermediate, and final evolutionary stages (or ‘types’ of individuals), respectively. The process initiates with a population of N asexually reproducing S_0 individuals. At every elementary time step, one individual is chosen to reproduce randomly with a probability that is proportional to its fitness. Another individual is chosen at random to die. The reproducing individual (‘mother’) gives rise to two daughters; one daughter replaces the mother, and the other replaces the dead individual, so

that the total population size remains constant at all times. When dividing, one daughter may mutate to the next evolutionary stage with probability u_1 if the mother is S_0 and u_2 if the mother is S_1 . No back mutation is permitted. The process continues until S_2 fixation is achieved or for T_{\max} elementary steps, whichever comes first.

Here, we refer to the simplest model where all individuals of the same evolutionary stage have the same fitness as the ‘zero-variance’ model. We compared this zero-variance model with a Moran model that incorporates fitness variation within evolutionary stages. In this new model, fitness variation within stages is caused by stochastic alterations in fitness that occur with every reproductive event. These alterations are described by a distribution of fitness effects that defines the probability of acquiring an alteration with a given fitness effect, and an effect lifetime τ that specifies the number of generations for which each alteration is inherited.

More specifically, let a particular individual in a population have absolute fitness w_i after its i th reproduction. This fitness

$$w_i = r_j \prod_{k=i-\tau+1}^i m_k$$

is the product of an evolutionary stage-specific fitness effect r_j for a stage- j individual and one or more stochastically acquired fitness alterations m_k . The values r_j (which have the value r_0 for initial-stage individuals, r_1 for intermediate-stage individuals, and r_2 for final-stage individuals; $r_2 > r_0 > r_1$) define the depth of the fitness valley. For simplicity, r_0 is set to 1 for the entire study. The fitness alterations m_k are drawn at each reproductive event k from a fitness effect distribution F with mean 1 and variance V . Mothers pass on the same value of the fitness alteration m_k to both of their daughters at reproduction k to recapitulate the high interdivision time correlations between sister cells observed in single cell experiments³. Each fitness alteration persists through exactly τ reproductive events, including the reproductive event that generated that alteration. Therefore, when $\tau = 1$, fitness alterations are not passed down to subsequent generations. In this case, during each reproduction event, a new fitness for the dividing cell is selected according to F , and the steady-state relative fitness distribution of a population of a single evolutionary is equivalent to F . Conversely, as τ goes to infinity, fitness alterations are permanent. The zero-variance model both represents the limit as τ goes to 0 and the limit as V goes to 0.

All fitness modifications are modeled on the multiplicative scale (i.e., as a percent change in fitness, rather than an addition or subtraction) to ensure that the relative fitness advantage or penalty associated with a mutation does not change as the absolute population mean fitness drifts upward in populations with long τ . However, our results also hold when the fitness modifications are modeled on an additive scale (Supplementary Figure 9); the disadvantage of this approach is that the fitness penalties and bonuses associated with a mutation diminish as the overall fitness of the population grows.

In most of our simulations, all sources of non-genetic fitness changes that occur during a single reproduction create a single fitness alteration with one combined fitness effect and a deterministic lifetime. However, in Supplementary Figure 7, each non-genetic fitness alteration that occurs during reproduction is modeled as a separate event. In this modified version of the model, a Poisson-distributed number of these events occurs during reproduction, with an average of one event occurring per reproduction. During subsequent generations, each individual alteration has a fixed probability of reversion, leading to fitness effect lifetimes that are geometrically distributed. Simulations of this evolutionary model are parameterized by the expected fitness effect lifetime τ .

Fitness effect distributions. We investigated four different distributions for F : the log-normal, double-exponential (Laplace), gamma, and centered Bernoulli distributions. Parameters for these distributions were chosen to provide mean 1 and the appropriate variance. The log-normal distribution was used for the majority of simulations, since the steady-state population fitness distribution is similar to the experimentally measured distributions, even when fitness effects do not persist beyond a single generation. The log-normal distribution provides similar results to simulations using the other distributions, including the heavier-tailed exponential and gamma distributions, which have previously been shown to describe mutational fitness effect distributions^{31,32}. The centered Bernoulli distribution represents the case in which all epigenetic changes have the same fitness effect magnitude and are beneficial or deleterious with the same probability, and is defined by the probability mass function

$$P(k) = \begin{cases} \frac{1}{2} & \text{for } k = -\sqrt{V} \\ \frac{1}{2} & \text{for } k = \sqrt{V} \end{cases}$$

Estimation of single cell fitness distributions. The CD8+ and L1210 fitness distributions were estimated from single cell interdivision time data for two types of murine cells⁴ (Fig. 1b). For the PC9 cell line, we leveraged previous work³³ demonstrating that an exponentially modified Gaussian provides a good fit to the intermitotic time distribution of these cells. We therefore used the maximum

likelihood parameter estimates for PC9 cells treated with dimethyl sulfoxide (DMSO) reported previously to estimate the PC9 fitness distribution²⁷ (Supplementary Figure 10). To estimate the corresponding fitness distributions for all three cell types, we considered fitness to be inversely proportional to interdive time; this assumption enables us to create an empirical fitness distribution for each cell type.

Comparison of simulated/experimental fitness distributions. To determine whether the experimentally measured CD8+ and L1210 single cell fitness distributions could have been generated by our evolutionary model with permanent non-genetic fitness alterations, we estimated the probability that the experimentally measured distributions would be at least as different from our simulated fitness distributions as we observed, using a two-tailed goodness-of-fit z-test³⁴. Specifically, we tested the null hypothesis that

$$H_0 : g(x) = h(x),$$

where $g(x)$ is the experimentally measured fitness distribution (CD8+ or L1210) and $h(x)$ is the simulated fitness distribution of our model, with a log-normal fitness effect distribution and non-genetic fitness effects that persist over 10 generations ($\tau = 10$). The experimental distribution is sampled by n i.i.d. measured cell fitnesses X_i , which are scaled and shifted to have the same mean and variance as the simulated distribution. The mean-squared error (MSE) between the distributions

$$\begin{aligned} \text{MSE} &= \int (g(x) - h(x))^2 dx \\ &= \int g^2(x) dx + \int h^2(x) dx - 2 \int g(x)h(x) dx \end{aligned}$$

can be estimated as

$$\widehat{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n \widehat{g}_{-i}(X_i) + \int \widehat{h}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \widehat{h}(X_i),$$

where $\widehat{g}_{-i}(x)$ is the leave-one-out Gaussian kernel density estimate of $g(x)$ from all experimental data except the measurement X_i and $\widehat{h}(x)$ is the Gaussian kernel density estimate of the simulated distribution $h(x)$. This estimated MSE is asymptotically normal under the null hypothesis, with the test statistic

$$\frac{\widehat{\text{MSE}}}{\left(\sum_{i=1}^n \sum_{j=1}^n K^2 \left(\frac{X_i - X_j}{h} \right) \right)^{1/2}} \xrightarrow{D} N(0, 1),$$

where K is the Gaussian kernel function with bandwidth h used to estimate $g(x)$ above.

Estimation of simulated tunneling and fixation probabilities. Our results focus on differences in the estimated final-stage fixation probability (i.e., proportion of simulations that result in fixation of final stage individuals before 3650 generations have passed) and the tunneling probability between simulated conditions. Here, we define the tunneling probability as the conditional probability that intermediate stage individuals fix at some point during the simulation, given that the final stage fixes by the end of the simulation. We estimated the tunneling probability as

$$\hat{P}(\text{tunnel}|\text{fix}) = \frac{\hat{P}(\text{tunnel}|\text{fix})}{\hat{P}(\text{fix})}$$

where the joint probability of tunneling and final stage fixation and the fixation probability are directly estimated from the proportion of simulations with final-stage fixation and tunneling or the frequency of final-stage fixation, respectively.

We used a similar strategy to estimate the conditional probability that a new intermediate-stage mutant will give rise to a final stage lineage that fixes by the end of the simulation, given that it arises with a specific relative fitness. Using Bayes' theorem, we estimated

$$\hat{P}(\text{fix final stage}|\text{fitness}) = \frac{\hat{P}(\text{fitness}|\text{fix final stage})\hat{P}(\text{fix final stage})}{\hat{P}(\text{fitness})}$$

by estimating the density of fitnesses of new intermediate-stage individuals that eventually lead to final-stage fixation (the conditional distribution $P(\text{fitness}|\text{fix final stage})$), the total density of all new intermediate stage individuals $P(\text{fitness})$, and the overall frequency of new intermediate stage individuals that eventually lead to final-stage fixation. In Fig. 4c, the fitness distributions for each condition were estimated from all intermediate individuals created by de novo mutation in 10,000 independent simulations (approximately 300,000 individuals total). Probability density functions were estimated by Gaussian kernel smoothing in R (version 1.0.143). The total probability of fixation of new intermediate-stage individuals was estimated as the proportion of all new intermediate individuals that generate a successful final stage lineage in 10,000 independent simulations per condition.

Study design and reproducibility. The number of simulations performed for each condition (typically 10,000 trials) was chosen such that enough adaptation events were observed to confidently estimate the mean tunneling probability for each condition. For a conservative estimate of a final-stage adaptation probability of 0.1, approximately 1000 adaptation events were observed, leading to an acceptable standard error of the estimated tunneling probability of at most 0.016 (CV ~2%).

The findings from this study were verified for limiting cases of the simulation parameters with separate custom simulation code by one of the authors who had not seen the software used to generate the data presented in the paper. None of the conditions tested had results that were unable to be reproduced. No data were excluded from analysis. As this was a theoretical study, no blinding or randomization methods were used.

Code availability. The stochastic evolutionary dynamics C++ software package used for all simulations is available on GitHub (https://github.com/Michorlab/evo_sim), with documentation. Input files to the simulation software used to generate the results presented here are available at <https://doi.org/10.7910/DVN/5D6YYP> within each individual data directory. Custom Python 2.7 and R scripts (R version 3.4.1) used for data analysis and visualization are available with the data at <https://doi.org/10.7910/DVN/5D6YYP>.

Data availability

Raw simulation data generated and analyzed in this study are available in the Harvard Dataverse at <https://doi.org/10.7910/DVN/5D6YYP>.

Received: 19 July 2018 Accepted: 28 August 2018

Published online: 26 September 2018

References

1. Perfeito, L., Fernandes, L., Mota, C. & Gordo, I. Adaptive mutations in bacteria: high rate and small effects. *Science* **317**, 813–815 (2007).
2. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
3. Sandler, O. et al. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519**, 468–471 (2015).
4. Kimmerling, R. J. et al. A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat. Commun.* **7**, 10220 (2016).
5. Keren, L. et al. Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* **166**, 1282–1294.e18 (2016).
6. Shaffer, S. A. et al. Memory sequencing reveals heritable single cell gene expression programs associated with distinct cellular behaviors. Preprint at <https://www.biorxiv.org/content/early/2018/07/27/379016> (2018).
7. Sigal, A. et al. Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
8. Sontag, L. B., Lorincz, M. C. & Georg Luebeck, E. Dynamics, stability and inheritance of somatic DNA methylation imprints. *J. Theor. Biol.* **242**, 890–899 (2006).
9. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
10. McVean, G. A. & Charlesworth, B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944 (2000).
11. Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647–661 (2011).
12. Neher, R. A. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu. Rev. Ecol. Evol. Syst.* **44**, 195–215 (2013).
13. Gillespie, J. H. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**, 909–919 (2000).
14. Neher, R. A., Shraiman, B. I. & Fisher, D. S. Rate of adaptation in large sexual populations. *Genetics* **184**, 467–481 (2010).
15. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
16. Levin, B. R., Perrot, V. & Walker, N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **154**, 985–997 (2000).
17. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
18. da Silva, J., Coetzer, M., Nedellec, R., Pastore, C. & Mosier, D. E. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* **185**, 293–303 (2010).
19. Nowak, M. A., Michor, F., Komarova, N. L. & Iwasa, Y. Evolutionary dynamics of tumor suppressor gene inactivation. *Proc. Natl Acad. Sci. USA.* **101**, 10635–10638 (2004).

20. Komarova, N. L., Sengupta, A. & Nowak, M. A. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J. Theor. Biol.* **223**, 433–450 (2003).
21. Iwasa, Y., Michor, F. & Nowak, M. A. Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579 (2004).
22. Haeno, H., Maruvka, Y. E., Iwasa, Y. & Michor, F. Stochastic tunneling of two mutations in a population of cancer cells. *PLoS ONE* **8**, e65724 (2013).
23. Weissman, D. B., Desai, M. M., Fisher, D. S. & Feldman, M. W. The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.* **75**, 286–300 (2009).
24. Neher, R. A. & Shraiman, B. I. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* **188**, 975–996 (2011).
25. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl Acad. Sci. USA* **108**, E67–E76 (2011).
26. Moran, P. A. P. *The Statistical Processes of Evolutionary Theory*. (Clarendon Press, London, 1962).
27. Leander, R., Allen, E. J., Garbett, S. P., Tyson, D. R. & Quaranta, V. Derivation and experimental comparison of cell-division probability densities. *J. Theor. Biol.* **359**, 129–135 (2014).
28. Beerenwinkel, N. et al. Genetic progression and the waiting time to cancer. *PLOS Comput. Biol.* **3**, e225 (2007).
29. Wloch, D. M., Szafraniec, K., Borts, R. H. & Korona, R. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* **159**, 441–452 (2001).
30. Potten, C. S., Kellett, M., Roberts, S. A., Rew, D. A. & Wilson, G. D. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* **33**, 71–78 (1992).
31. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
32. Kassen, R. & Bataillon, T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* **38**, 484–488 (2006).
33. Golubev, A. Exponentially modified Gaussian (EMG) relevance to distributions related to cell proliferation and differentiation. *J. Theor. Biol.* **262**, 257–266 (2010).
34. Fan, Y. Testing the goodness of fit of a parametric density function by kernel method. *Econom. Theory* **10**, 316–356 (1994).

Acknowledgements

Portions of this research were conducted on the O2 High Performance Compute Cluster, supported by the Research Computing Group, at Harvard Medical School. See <http://rc.hms.harvard.edu> for more information.

Author contributions

F.M. conceived of the idea. D.V.E. and T.M. performed the analyses. D.V.E. produced all figures. All authors wrote the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s42003-018-0160-1>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018