PAPER

Biophysical constraints determine the selection of phenotypic fluctuations during directed evolution

IOP Publishing Phys. Biol. 15 (2018) 065003 https://doi.org/10.1088/1478-3975/aac4e6

Physical Biology

PAPER

Biophysical constraints determine the selection of phenotypic fluctuations during directed evolution

Hong-Yan Shih[1], Harry Mickalide[1], David T Fraebel[1], Nigel Goldenfeld[1,2] and Seppe Kuehn[1,2]

[1] Department of Physics and Center for the Physics of Living Cells, Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1110 West Green St., Urbana, IL 61801, United States of America

[2] Author to whom any correspondence should be addressed.

E-mail: nigel@uiuc.edu **and** seppe@illinois.edu

RECEIVED
5 October 2017

REVISED
18 April 2018

ACCEPTED FOR PUBLICATION
15 May 2018

PUBLISHED
17 July 2018

Keywords: phenotypic fluctuations, directed evolution, chemotaxis, Monte Carlo simulations

Supplementary material for this article is available online

Abstract

Phenotypes of individuals in a population of organisms are not fixed. Phenotypic fluctuations, which describe temporal variation of the phenotype of an individual or individual-to-individual variation across a population, are present in populations from microbes to higher animals. Phenotypic fluctuations can provide a basis for adaptation and be the target of selection. Here we present a theoretical and experimental investigation of the fate of phenotypic fluctuations in directed evolution experiments where phenotypes are subject to constraints. We show that selecting bacterial populations for fast migration through a porous environment drives a reduction in cell-to-cell variation across the population. Using sequencing and genetic engineering we study the genetic basis for this reduction in phenotypic fluctuations. We study the generality of this reduction by developing a simple, abstracted, numerical simulation model of the evolution of phenotypic fluctuations subject to constraints. Using this model we find that strong and weak selection generally lead respectively to increasing or decreasing cell-to-cell variation as a result of a bound on the selected phenotype under a wide range of parameters. However, other behaviors are also possible, and we describe the outcome of selection simulations for different model parameters and suggest future experiments. We analyze the mechanism of the observed reduction of phenotypic fluctuations in our experimental system, discuss the relevance of our abstract model to the experiment and explore its broader implications for evolution.

1. Introduction

Natural selection acts at the level of the phenotype. Unlike genomes, phenotypes can be highly variable over the lifetime of a single organism or heterogeneous across a genetically identical population. Given the central role of the phenotype in selection, phenotypic fluctuations are believed to play an important role in evolution.

Therefore, understanding the evolutionary origins and impacts of phenotypic fluctuations will be central to any quantitative theory of evolution. Environmental factors provide selection pressure that prefers certain phenotypes, through which the mutant genotypes that represent similar phenotypes can be selected. Phenotypic fluctuations can arise by stochastic variation in gene expression [1], which can be associated with physiological responses to environmental variation (plasticity) [2]. In bacteria, non-genetic phenotypic variability in a population is critical for survival in the presence of antibiotics [3]. Further, non-genetic variation is present in bacterial swimming behavior [4] and is thought to be adaptive [5].

The role of phenotypic fluctuations in evolution, and how genetic variation alters phenotypic fluctuations, has been the subject of theoretical and experimental investigations since Baldwin [6]. Waddington presented compelling arguments for the role of phenotypic plasticity in facilitating evolution through genetic assimilation [7], and conceptual models of this effect abound [8]. Notably, Sato *et al* formulated a phenomenological model based on the fluctuation-dissipation theorem, which postulates that phenotypes exhibiting larger fluctuations should evolve more rapidly under selection [9]. The theory was tested in a directed evolution experiment by constructing a diverse popula-

tion of green fluorescent protein (GFP) expressing *Escherichia coli* mutants synthetically and then selecting for higher levels of GFP fluorescence. The study showed that directed selection for increasing mean fluorescence resulted in reduced cell-to-cell variability in fluorescence intensity [9]. Conversely, a subsequent series of experimental studies showed that strong selection on the phenotype led to an increase in phenotypic fluctuations [10]. The interpretation of this experiment is complicated, however, because there were only a few clones in the system, and the population seemed to split into two types of mutant distinguished by the variance in their phenotype fluctuations [10]. Similarly, in directed evolution experiments of cell size in *E. coli* a decrease of cell-to-cell variation in size was reported for weak selection whereas little change in cell size fluctuation was observed under strong selection [11].

Phenotypes arise from genotypes through the processes of transcription and translation. Therefore, any generic features of the evolution of phenotypic fluctuations might be illuminated by considering universal aspects of gene expression. Protein copy number distributions have been measured in a variety of microbial species, for example in cultured populations of bacteria [12–14] and yeast [13, 14] and in single-cells [15–17]. These studies show that the probability density of protein copy number across a population is consistently non-Gaussian and highly skewed, and reportedly well fit by gamma [15], extreme value (Fisher–Tippett–Gumbel [14] or Frechet [13]) or log-normal [12, 17] distributions, all of which are similar in shape. Regardless of the precise form of the distribution reported, one trend is clear: the standard deviation σ is a monotonically increasing function of the mean, and the distributions can be collapsed onto a single universal curve using reduced coordinates $(n - \langle n \rangle)/\sigma$ [13, 17]. If a phenotype can be associated with a particular dominant protein, then as the phenotype and hence the protein copy number is increased during a directed evolution experiment, one might naively expect the phenotypic variation to increase as well, a result that is not generically found to be true. In reality, the relationship between protein copy number and phenotype is more complex, reflecting regulation, inhibition, and feedback. Therefore, the precise relationship between protein copy number and phenotype remains unclear, with little likelihood of a universal connection, even if the global statistics exhibit universal functional forms.

Direct empirical evidence for the relationship between phenotypic fluctuations and long-term evolution remains limited. Notable exceptions include retrospective studies of hemoglobin binding affinity across mammals [18], but even this study does not make direct measurements of phenotypic fluctuations in time or across individuals. While experimental evolution has revealed striking examples of phenotypic evolution [19–22], quantitative measurements of phenotypic fluctuations in many of these experiments have not been made. As a result, conceptual or quantitative models of the evolution of phenotypic fluctuations remain untested.

Here we present a joint theoretical and experimental investigation of how phenotypic fluctuations evolve under selection. We use high-throughput phenotyping to show that the phenotypic variation in the population declines when bacteria are selected for faster migration through a porous environment [22]. We then present a simple model of directed evolution which allows us to interrogate how selection strength and mutations result in the evolution of phenotypic fluctuations. We show that, depending on the strength of selection, phenotypic fluctuations can decline when phenotypes are subjected to constraints even when there is no mechanistic link between the mean trait value and phenotypic fluctuations. We discuss the relevance of this theoretical result to the experimentally observed reduction in phenotypic fluctuations. Finally, we discuss the possible biological mechanisms underlying the experimentally observed reduction in phenotypic fluctuations in the context of our model.

2. Evolution of faster migration in *E. coli*

Growing populations of motile, chemotactic bacteria migrate outward when inoculated into a soft agar plate containing growth medium and a chemoattractant [23, 24]. As cells swim and divide in this porous environment local depletion of nutrients establishes a spatial nutrient gradient which drives chemotaxis through the three-dimensional agar matrix and subsequent nutrient consumption. Microscopically, cells move through the porous environment by executing runs, at a speed $|v_r| \sim 20$ μm s^{-1} for a run duration $\tau_r \sim 1$ s, and tumbles which rapidly reorient the cell in $\tau_t \sim 0.1$ s. Tumbles are essential for avoiding obstacles in order to successfully navigate the soft agar [23]. The result is a macroscopic colony that expands radially through the bulk of the plate at a constant speed after an initial growth phase. We selected populations of *E. coli* (MG1655-motile, Coli Genetic Stock Center, Yale University #6300) for faster migration through soft agar by repeatedly allowing a colony to expand for a fixed interval, sampling a small population of cells from its outer edge and using a portion of this sample to inoculate a new plate while preserving the remainder cryogenically (figure 1). In rich medium conditions (LB, 0.3% w/v agar, 30 °C), we sampled after 12 h of expansion for a total of 15 rounds of selection. By performing time-lapse imaging on the expanding colonies, we found that the migration rate approximately doubled over the first five rounds of selection and continued to increase marginally in subsequent rounds. We found that this increase was reproducible across replicate experiments.

To understand the mechanism by which faster migration evolved, we performed single cell tracking

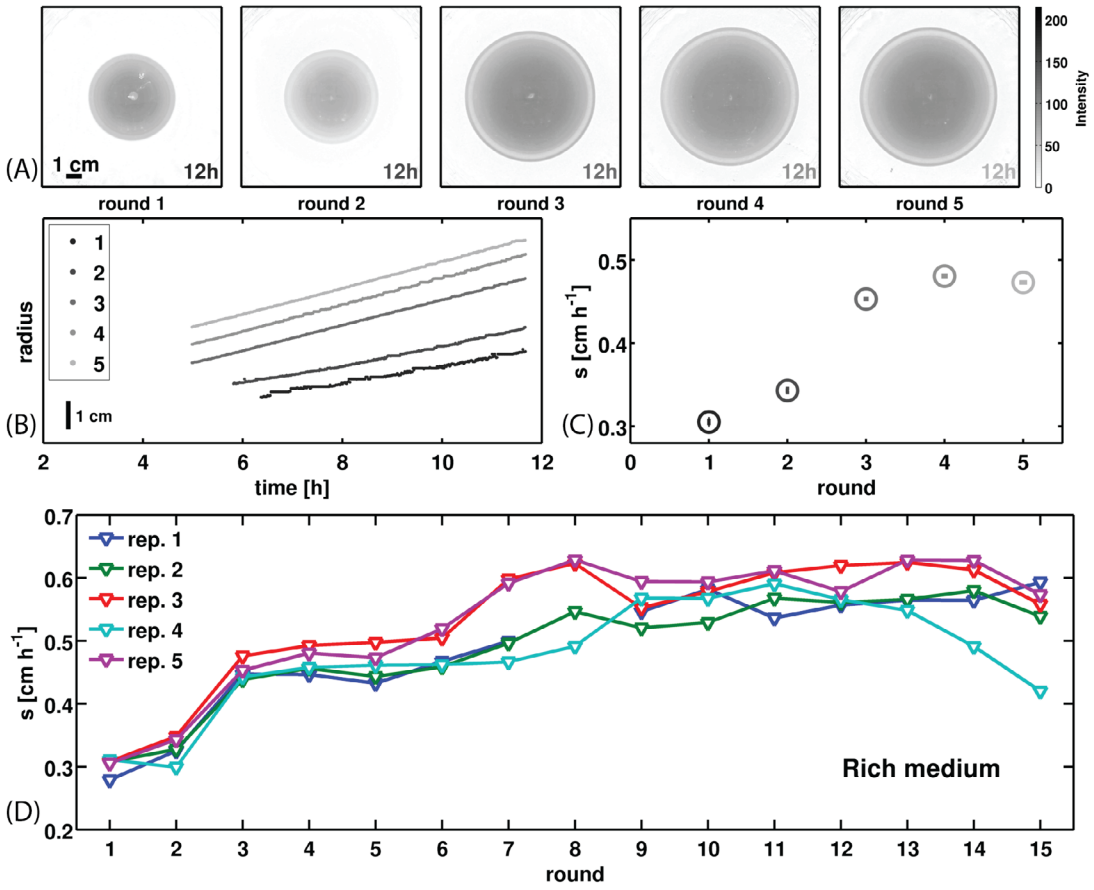

Figure 1. Selection for faster bacterial migration: (A) shows images of *E. coli* colonies in low viscosity agar plates after 12 h of expansion. After 12 h a sample of the outer band of cells is taken and approximately 6 cells are used to initiate another identical agar plate (second panel). This process is repeated for 15 rounds of selection where a round consists of colony expansion in a single plate. The color bar to the right applies to all panels, with darker gray indicating higher cell density. The scale bar in the left panel applies to all panels in (A). (B) The radius of each colony in (A) as a function of time, lighter shades of gray denote later rounds of selection and correspond to labels in (A). Traces are offset vertically for clarity, note scale bar lower left. (C) The rate of the linear portion of the colony expansion as a function of the round of selection for the plates shown in (A)–(B). (D) The evolutionary process outlined in (A)–(C) was carried out in five independent experiments. Each line corresponds to an independent selection experiment. Round 8 for replicate 1 is missing due to failure of the imaging device. The data in panels (A)–(C) are from replicate 5. Errors in rate of expansion are smaller than the size of markers. Data recapitulated from [22].

on hundreds of individuals from the ancestral strain as well as from strains isolated after 5, 10 and 15 rounds of selection. Individual cells were trapped in a circular microfluidic chamber in the same medium in which the selection was performed and recorded while swimming for 5 min per cell. Swimming cells were imaged at 30 Hz, automated tracking routines constructed swimming trajectories from these movies and runs and tumbles were automatically identified as described previously [22]. This measurement permitted us to capture the swimming behavior of hundreds of single bacterial cells in the absence of chemical gradients. We found that the average run speed increased by approximately 50% during selection, while the duration of run and tumble events declined (figure 2). The maximum growth rate, which was measured in a separate experiment by monitoring the optical density of a well-stirred liquid culture declined over the course of selection. The trade-off between swimming speed and growth rate is the subject of a separate study [22] and similar trade-offs have been observed elsewhere [25].

3. Phenotypic fluctuations decline with selection

Phenotypic fluctuations have previously been characterized in several ways. In some cases, fluctuations refer to the time-dependence of a specific phenotypic parameter during the lifetime of an individual [26]. In other studies, fluctuations refer to cell-to-cell variation in time-averaged phenotypic measurements over a population [4, 27–29]. Here we use the latter approach, which is shown schematically in figure 3. Briefly, from run-tumble events performed by each individual cell we computed an average phenotype (run duration, tumble duration, and run speed) for each cell. From these data we computed a distribution of average phenotypes across individuals in the population, and thus extracted the standard deviation over the population of a given phenotype. This standard deviation directly measures cell-to-cell variation, as sketched in figure 3.

To define phenotypic fluctuations more explicitly consider a single *E. coli* cell which exhibits a series of

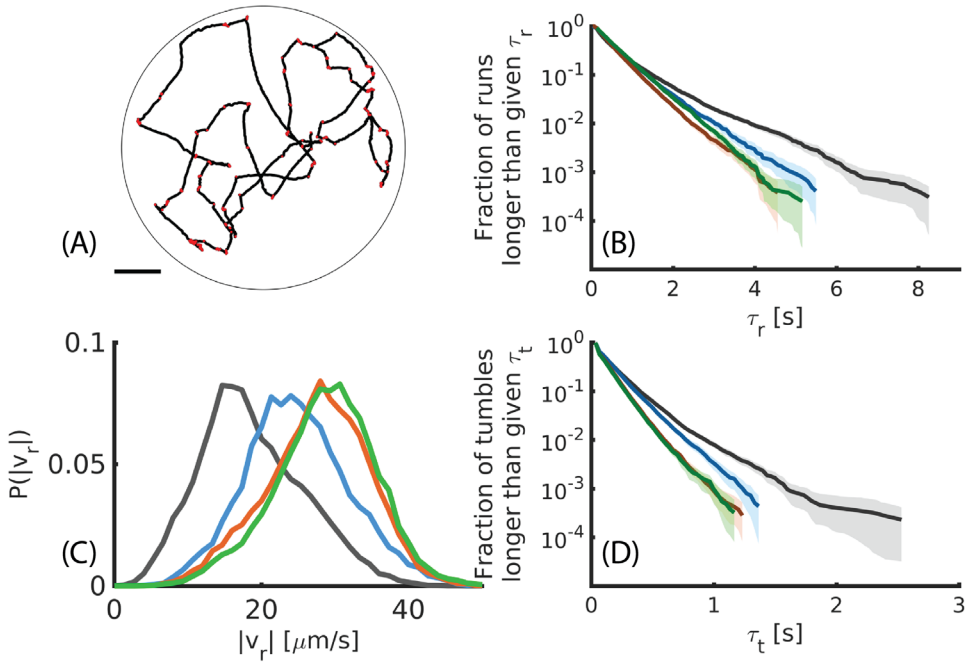

Figure 2. Dynamics of phenotypic evolution: (A) an example 50 s long swimming trajectory for a single cell trapped in a microfluidic chamber. The boundary of the chamber is shown by the light black circle. Running events are shown in black and tumble events in red. Scale bar is 50 μm. (B) Aggregate complementary cumulative distribution functions of run durations observed from cells isolated prior to selection (founder,black) and after 5 (blue), 10 (orange) and 15 (green) rounds of selection. Strains tracked were isolated from replicate 1 in figure 1. Distributions are constructed from all run events that were not interrupted by collisions with the chamber boundary for 140 (founder), 79 (round 5), 97 (round 10) and 96 (round 15) individuals executing a total of 19597, 12217, 18505 and 15928 run events respectively. The mean and standard deviation of run durations are (mean:sd) 0.66 s:0.78 s, 0.63 s:0.61 s, 0.58 s:0.51 s and 0.64 s:0.57 s respectively. Shaded regions indicate 95% confidence intervals from bootstrapping. (C) Distributions of run speeds ($|v_r|$) for the four strains shown in (B), colors from (B) apply. Distributions are constructed by computing an average speed for each run event. Means of these distributions are 18.7 μm s^{-1} (founder), 24.9 μm s^{-1} (round 5), 27.6 μm s^{-1} (round 10), and 28.7 μm s^{-1} (round 15). The increase in $|v_r|$ is statistically significant between each successive population ($p < 0.001$, rank sum test). (D) shows the tumble duration distributions for the same four strains shown in panels (B)–(C). The mean and standard deviation of tumble durations are (mean:sd) 0.18 s:0.20 s, 0.17 s:0.16 s, 0.14 s:0.13 s and 0.14 s:0.12 s for founder, round 5, 10 and 15 respectively. Data are recapitulated from [22].

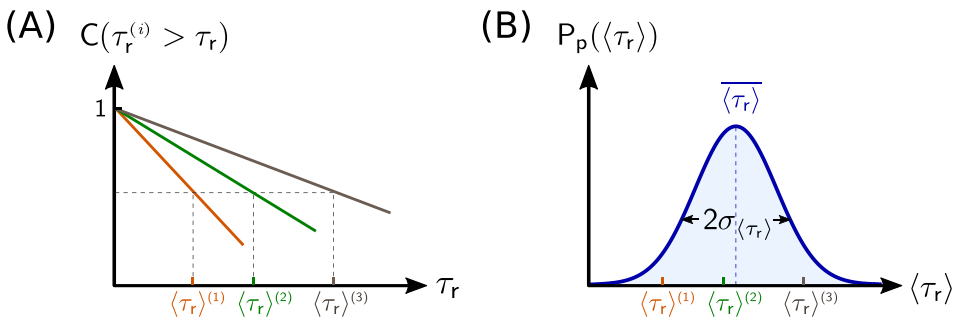

Figure 3. Illustration of phenotype distributions: (A) An example of the complementary cumulative distribution function for run duration from statistics of all run events of different individuals. The average run duration of individual i ($\langle\tau\rangle^{(i)}$) is read off by fitting the exponential distribution. (B) Distribution of individual run duration $\langle\tau\rangle^{(i)}$ in the population generated from (A). The cell-to-cell variation is characterized by the standard deviation of $P_p(\langle\tau_r\rangle)$, $\sigma_{\langle\tau_r\rangle}$.

runs and tumbles. Each run event is described by a run duration (τ_r) and a run speed ($|v_r|$) and each tumble by a tumble duration (τ_t) and an angular velocity (ω_t). Even in an unstimulated environment where no gradients are present, τ_r will vary between run events, and the distribution exhibited by individual i is given by $P(\tau_r^{(i)})$. Each run event for this individual has a duration drawn from this distribution. Similar distributions exist for $|v_r|$, τ_t, and ω_t, but ω_t is difficult to measure accurately for single cells, and we omit this parameter from consideration. We consider the phenotype of a single cell to be the mean of these distributions. Thus a complete description of unstimulated swimming behavior of a single cell is captured by the set of phenotypes $\chi^{(i)} \in \{\langle\tau_r\rangle^{(i)}, \langle|v_r|\rangle^{(i)}, \langle\tau_t\rangle^{(i)}\}$, where $\langle\cdot\rangle^{(i)}$ denotes an average over all events exhibited by individual i. In a population, phenotypic traits can be described by a distribution $P_p(\chi)$ that governs

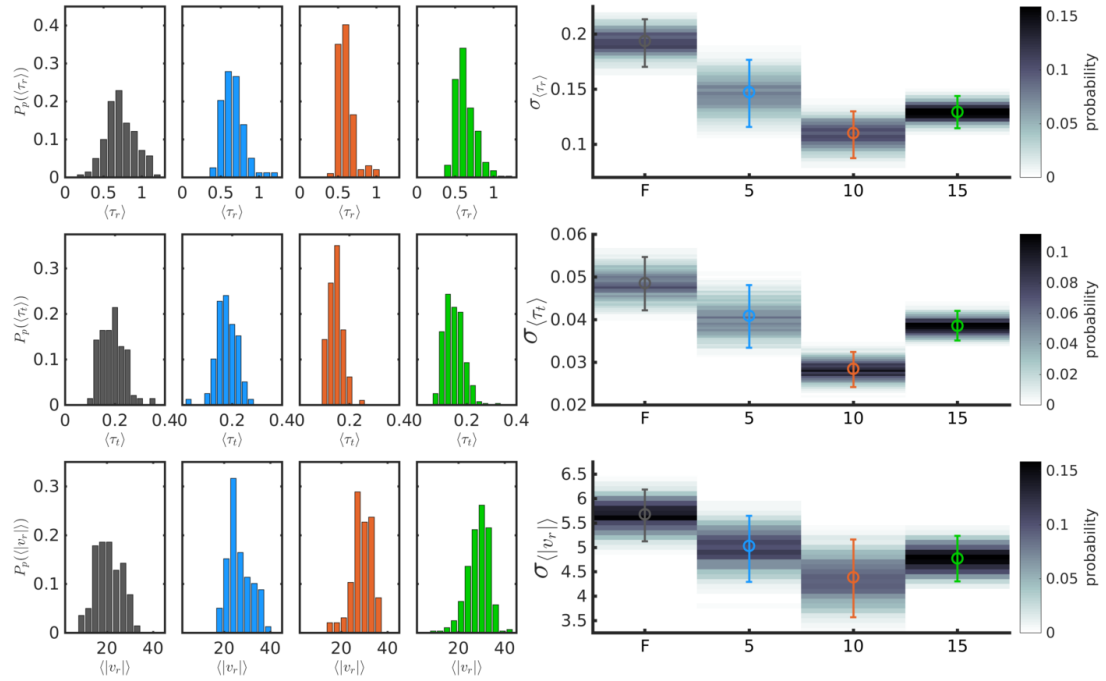

Figure 4. Cell-to-cell behavioral variation declines with selection: phenotype distributions ($P_p(\chi)$) for (A) τ_r (B) τ_t and (C) $|v_r|$ from raw data. Individuality for evolved populations for (D) τ_r (E) τ_t and (F) $|v_r|$. We compute $\langle\tau_r\rangle$, $\langle\tau_t\rangle$ and $\langle|v_r|\rangle$ for each individual tracked and a standard deviation across individuals for each parameter (σ_χ). σ_χ was computed for 140 cells (founder), 79 cells (round 5), 97 cells (round 10) and 96 cells (round 15). The circles show the sample σ_χ for each population. 95% confidence intervals from bootstrapping for each population are given by the error bars. Colormap shows the probability distribution of σ_χ from bootstrapping. Note distinct colorbars for each panel. All populations exhibit a decline in σ_χ relative to founder that is significant ($p < 0.05$, permutation test) except for $\sigma_{\langle\tau_t\rangle}$ and $\sigma_{\langle|v_r|\rangle}$ in round 5. The increase in σ_χ between rounds 10 and 15 is only statistically significant for tumble duration ($p = 5 \times 10^{-4}$).

the probability that an individual has a specific value for each trait χ. The distributions $P_p(\chi)$ for the founding strain and individuals isolated after 5, 10 and 15 rounds of selection are shown in figures 4(A)–(C). We quantify phenotypic fluctuations, or cell-to-cell variation, by the standard deviation across the population in each trait, for N cells this is computed as: $\sigma_\chi = \sqrt{\frac{1}{N}\sum_i(\chi^{(i)} - \langle\chi^{(i)}\rangle)^2}$. We note that σ_χ describes phenotypic variation driven by both genetic and non-genetic variation in the population except in cases of clonal populations, where σ_χ is due to non-genetic effects alone.

To experimentally quantify phenotypic fluctuations we computed average run durations, tumble durations and run speeds on a per cell basis. Explicitly, if cell i executes M runs during the 5 minutes of tracking we compute $\langle\tau_r\rangle^{(i)} = \frac{1}{M}\sum_{j=1}^{M}\tau_{r,j}$. To quantify the cell-to-cell variation we then compute the standard deviation across individuals $\sigma_{\langle\tau_r\rangle}$. We compute identical statistics for the tumble duration τ_t and the run speed $|v_r|$ for founding populations and populations isolated after 5, 10 and 15 rounds of selection. Figures 4(D)–(F) shows the standard deviations across the population (σ_χ) for $\chi \in \{\langle\tau_r\rangle, \langle\tau_t\rangle, \langle|v_r|\rangle\}$, indicating a significant decline in the cell-to-cell variation during selection. In particular, we observe a significant decline between founding population and rounds 10 and 15 for all phenotypic parameters. We conclude that selection for faster migration results in reduced phenotypic fluctuations in the population.

The common interpretation for the utility of phenotypic variation is that it may increase survival probability under environmental changes by providing variation with every generation as opposed to genetic mutations which occurs less frequently [3, 5]. Whether populations are shaped more by phenotypic variation or genotypic variation depends on the degree of phenotypic variation and on the strength and types of environmental selection. Is this reduction a special feature of the experiment, or can it be understood from general principles? To address this, we describe below an abstract computational model which is independent of the mechanistic details of our particular experiment. We ask how the process of iterated selection, whereby cells from the tail of a phenotypic distribution are propagated to the next round, alters cell-to-cell variation. Our goal with the simulation is to predict how the evolution of cell-to-cell variation depends on the strength of selection.

4. Abstract model of directed evolution of phenotypic fluctuations

The genotype-phenotype map determines the phenotype of an organism with a given genotype. How phenotypic selection is coupled to genetic variation is an important question whose answer illuminates

fundamental questions such as the evolutionary rate and the evolvability of organisms. In general, this mapping is a multi-dimensional function that is governed by complex biological features such as gene regulatory and metabolic networks. As such, in laboratory-based directed evolution experiments the evolutionary dynamics of a specific phenotype are difficult to understand in terms of genetic variation alone. Therefore, we seek a framework that does not rely on an explicitly modeled mapping from genotypes to phenotypes. For simplicity, we present a computational model of adaptation of a single effective phenotype and its associated genotype, representing a projection of a multi-dimensional phenotype/genotype evolving under selection. The idea is related to previous population genetics models [30], but instead of assuming continuous selection due to an assumed fitness landscape, we specify selection through a population bottleneck that is decoupled from the rate of growth. We use this model to calculate the evolution of phenotypic variation under selection. The model is necessarily stochastic in order to capture the dynamics of fluctuations. We do not specify any explicit mechanism for genotype-phenotype mapping or how its functional form changes during evolution. Instead, phenotypes are random numbers generated from a Gaussian mapping function whose mean is identified with a genotype and whose variance reflects phenotypic fluctuations across individuals with that genotype. The mean and variance change in evolutionary processes such as point mutations. Contrary to the conventional population genetics argument for directed evolution that predicts decrease in the variance of phenotype as a result of avoiding deviation from the peak in the fitness landscape, we attempt to understand how the various factors can affect the evolutionary trajectory in a minimal and general model.

4.1. Main features of the abstract model

Our abstract model captures key features of a fully realistic model built on a lower-level description such as gene expression. The main experimentally-relevant factors considered in this abstract model include strength of bottleneck selection, mother-daughter correlation and mutations. The mother-daughter correlation (or epigenetic inheritance) describes the degree of gene expression level that is passed on to descendants in the absence of mutations and determines how well preserved a phenotype is in subsequent generations. Mutations stochastically induce changes in the phenotype (χ). We focus on the effect of the strength of bottleneck selection and the mutation rate. The correlation between mother and daughter is effective in accumulating phenotype changes during directed evolution if the correlation is high. However, measurements showed that the mother-daughter correlation is around 0.5 (data shown in figure S1 (stacks.iop.org/PhysBio/15/065003/mmedia)) and is the same for both the founder and the evolved strains. Moreover, it was shown in the relaxation experiment in [22] that after about 140 generations of growth in well-mixed liquid conditions no additional mutations occurred and the fast migrating phenotype was retained throughout this extended growth period. Therefore, we set the mother-daughter correlation to not evolve in the directed evolution in our model. Finally, the bottleneck selection is applied in the trait space instead of the real space, and therefore the details of the experiment including the consumption of nutrients and the process of chemotaxis are not explicitly represented in the model in order to reduce complexity.

In addition, traits such as run speed cannot physically evolve to infinitely large values and thus should be bounded by a threshold χ_c. The threshold on phenotype represents a limitation of the corresponding cellular machinery, and therefore it fluctuates between cells in general. Due to the threshold, the trait χ converges when the mean of phenotype of the population gets closer to the threshold, and therefore the effective evolutionary rate decreases. We also anticipate that the convergence to the threshold can lead to skewness in the phenotype distribution, because fluctuations in the phenotype cannot exceed the threshold.

4.2. Naïve prediction for the effects of variation in selection strength

We expect that one of the relevant control parameters is the strength of the population bottleneck selection. We note that in the simulation multiple genotypes can coexist in the population at variable frequencies. Intuitively, without any physically-determined threshold on phenotype χ, individuals who evolve higher mean trait value and larger phenotypic fluctuations of their genotypes are expected to preferentially populate the right-most tail of the population trait distribution, and so will have a higher probability to be selected. Therefore, after population amplification where selection is absent, the overall phenotypic variance in the population would be expected to increase monotonically. However, once the population trait distribution approaches the threshold, the mean trait value of the genotypes of the selected individuals gets close to the threshold, and mutants with similar mean but different phenotypic fluctuations can arise. If the selection strength is strong, genotypes with both large or small isogenic fluctuations can both contribute large phenotype values near the threshold and be selected, so that the phenotypic variation would be expected to increase. On the other hand, genotypes with large isogenic fluctuations will have significant weighting at smaller χ, and therefore if selection strength is not strong enough, genotypes with smaller isogenic fluctuations are more likely to be selected, potentially leading to a decrease in the overall variance.

These naïve and intuitive arguments, however, do not account for the effects of individual variations in threshold, variations of threshold from generation to generation, and the effects of mutations. Our simulation results, described below, reveal a more subtle and complex series of outcomes in the evolution of phenotypic fluctuations. As a result, accurately predicting the dynamics requires stochastic quantitative models, and cannot be reliably carried out with naïve arguments.

4.3. Detailed description of the abstract model

In our abstract model, each individual i is represented by a random phenotype value $\chi^{(i)}$ which is determined by the individual's genotype g. $\chi^{(i)}$ is generated from a normal distribution $P(\chi)$ whose mean is $\mu_\chi(g)$ and whose variance is $s^2_\chi(g)$ in the absence of mother-daughter correlations. This abstracted phenotype is intended to represent any observable phenotypic variable. We assume that the phenotype does not change within the individual's lifetime. In our abstract model of directed evolution, the phenotypic trait χ is not explicitly stipulated. Instead, our abstract model is intended to explore the dynamics of phenotypic evolution under generic assumptions about how traits are passed between generations and respond to mutations.

Individuals reproduce and the offspring acquire mutations with probability ν, causing the daughter's genotype g' to be distinct from the mother's (g). Therefore, the daughter's phenotype follows another normal distribution with distinct mean $\mu_\chi(g')$ and distinct variance $s^2_\chi(g')$.

In the absence of mutations (i.e. within a clonal population derived from a single genotype g), the phenotypes of each new cell are generated based on a bivariate gaussian distribution $P(\chi^{(i)}, \chi^{(i')})$ with mother-daughter correlation coefficient ρ that captures the fact that daughter cells have phenotypes $\chi^{(i')}$ which is correlated with those of their mother $\chi^{(i)}$ [31]. Phenotypic correlations between generations in clonal populations can arise from protein copy number fluctuations or non-genetic changes in gene expression [32, 33]. For an individual i', which results from fission of individual i, its phenotype $\chi^{(i')}$ follows the conditional distribution of the variable $\chi^{(i')}$, given a known value of $\chi^{(i)}$ [31]:

$$P(\chi^{(i')}|\chi^{(i)}) \sim \mathcal{N}(\mu_\chi(g) + \rho(\chi^{(i)} - \mu_\chi(g)), (1-\rho^2)s^2_\chi(g)), \quad (1)$$

where $\mathcal{N}(\mu, s^2)$ is a normal distribution with mean μ and variance s^2.

We calculate these dynamics, along with the procedure for directed evolution through selection, as follows:

(i) N_s individuals from a single genotype $g = g_0$ are generated from $P(\chi) = \mathcal{N}(\mu_\chi(g_0), s^2_\chi(g_0))$, as illustrated in figure 5(A). These N_s clonal individuals are defined as the founder strain, which by construction is a population with a normal distribution of different phenotypes.

(ii) Each individual with phenotype $\chi^{(i)}$ creates a new individual with phenotype $\chi^{(i')}$. The new individual mutates to a new genotype $g = g_1 \neq g_0$ with a rate ν:
   (a) If it mutates, $\chi^{(i')}$ is generated from $P(\chi) = \mathcal{N}(\mu_\chi(g_1), s^2_\chi(g_1))$, where $\mu_\chi(g_1)$ and $s_\chi(g_1)$ are generated from $\mathcal{N}(\mu_\chi(g_0), \eta^2_{\mu_\chi})$ and $\mathcal{N}(s_\chi(g_0), \eta^2_{s_\chi})$ respectively. The variances $\eta^2_{\mu_\chi}$ and $\eta^2_{s_\chi}$ are assumed to be constant for all parent genotypes (g_0).
   (b) If the new individual does not mutate, $\chi^{(i')}$ updates based on equation (1).An example of the relationship between different phenotypes and the reproduction process is shown in figures 5(B) and (C). During reproduction we neglect the degradation of individuals, and thus the population doubles after one generation. Each individual in the doubled population generates a new individual in the next generation following step (ii)a or (ii)b. We assume that the mother-daughter correlation (ρ) does not evolve. After m generations, selection is applied to the whole population with $N_f = N_s \times 2^m$ individuals.

(iii) To apply selection, N_r individuals with the largest χ values are chosen from the population. The selection fraction N_r/N_f is defined to be the selection strength. N_s individuals are further randomly selected from the N_r individuals to be the seed population for the next round. N_r is analogous to the outer edge population sampled with a pipette in the experiments, and N_s represents the individuals that are used to inoculate the new plate. In experiments, $N_f \sim 10^{10}$, $N_r \sim 10^8$ and $N_s \sim 10^6$. Thus, in the bacterial chemotaxis experiments $N_f \gg N_r \gg N_s$.

(iv) In the new round, step (ii) and (iii) are repeated for the N_s individuals from the previous round.

(v) The phenotypic variance in N_s individuals at the end of each round is measured by growing a population to $N_l = N_s \times 2^l$ individuals by repeating step (ii)b without mutations. This mimics the experimental process of single cell tracking in liquid media, where populations are amplified by growth in well-mixed liquid conditions and presumably mutations can be neglected.

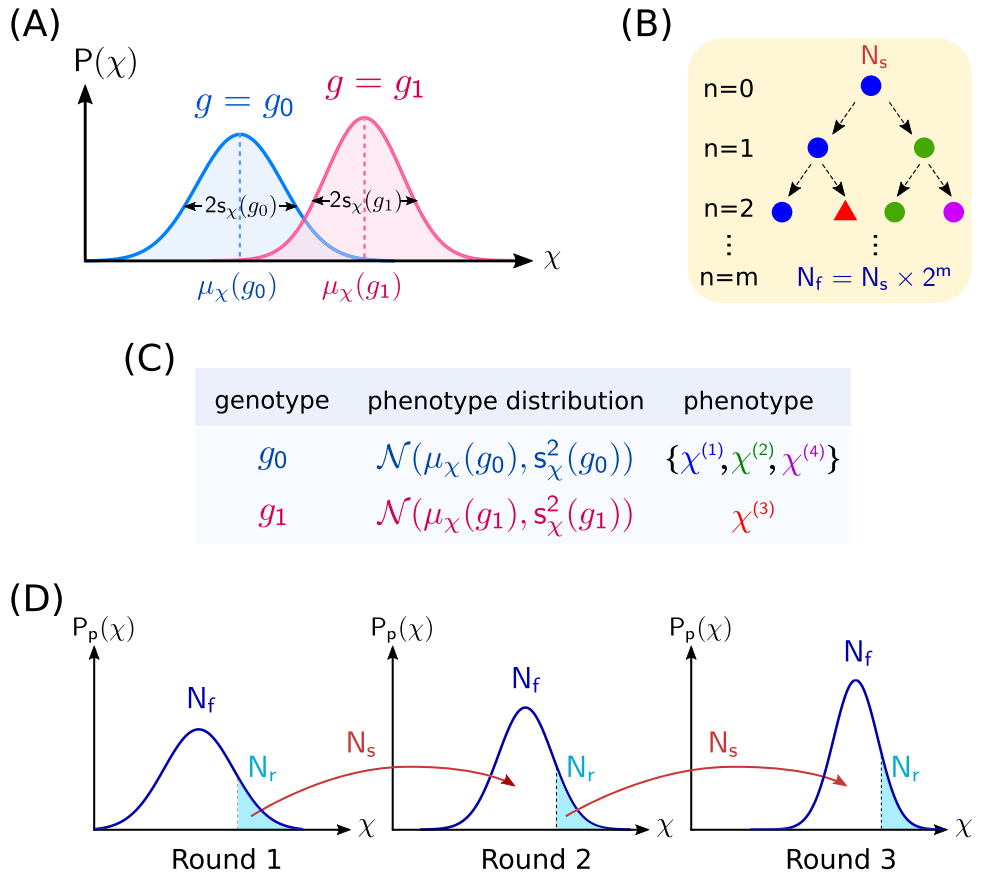

Figure 5. Scheme of the abstract model. Illustration of selection procedure (see text for definition of notation): (A) phenotype distributions for two genotypes (g_0, g_1). The phenotype of each genotype g_i is described by a normal distribution with mean $\mu_\chi(g_i)$ and standard deviation $s_\chi(g_i)$. (B) Initially the founder strain with N_s individuals whose phenotypes $\chi^{(i)}$ are drawn from $\mathcal{N}(\mu_\chi(g = g_0), s^2_\chi(g = g_0))$ is generated. The N_s individuals reproduce new individuals in the first round with a mutation rate ν. For example, for one of the initial N_s individuals with the founder genotype g_0 (circle) and a certain phenotype ($\chi^{(1)}$, in blue color) which is determined by mother–daughter correlation based on equation (1), its daughter may have the same genotype but different phenotype ($\chi^{(2)}$, in green color) if it does not mutate. If the daughter mutates, the daughter is assigned a new genotype (triangle) with $\mu_\chi(g_1)$ from $\mathcal{N}(\mu_\chi(g_0), \eta^2_{\mu_\chi})$ and $s_\chi(g_1)$ from $\mathcal{N}(s_\chi(g_0), \eta^2_{s_\chi})$, and its phenotype $\chi^{(3)}$ is drawn from $\mathcal{N}(\mu_\chi(g_1), s^2_\chi(g_1))$ (shown in red). All genotype and phenotype values are truncated by a random upper bound χ_c chosen from $\mathcal{N}(\mu_{\chi_c}, \eta^2_{\chi_c})$. (C) shows a table of phenotypes (χ^i) and their corresponding genotypes and phenotype distributions. Note that individuals with the same genotype stochastically differ in their phenotypes (first row). After m generations of the process shown in (B), the population becomes $N_f = N_s \times 2^m$. (D) The top N_r individuals are selected from N_f individuals, and N_s individuals are randomly sampled from N_r individuals to start the second round. In the next round, N_s individuals repeat reproduction steps in (A) until the population reaches N_f again. How close the average phenotype of N_f in the next round is to the average phenotype of N_r of the previous round depends on how small m is and how high the correlation between mother and daughter is. At the end of each round, the selected N_s individuals reproduce for l generations without mutations. These $N_l = N_s \times 2^l$ individuals represent the population of each strain grown in liquid media prior to single-cell tracking.

The parameters in the simulations are: $N_s = 100, m = 10, l = 10, \mu_\chi(g_0) = 40, s_\chi(g_0) = 8, \eta_{\mu_\chi} = 3, \eta_{s_\chi} = 1$, with $\nu = 0.2$. Simulations were run over 120 rounds. The stochastic values of χ, μ_χ and s_χ are binned to create finite differences between trait values. The bin sizes in the simulations are 1, 3 and 1 respectively. The selection process is described in figure 5(D).

These simulations do not directly stipulate how the phenotypic fluctuations within a given genotype $s_\chi(g)$ evolve—e.g. these can increase or decrease relative to the parent genotype g. This is intended to avoid any bias on phenotypic fluctuations with respect to the evolving mean trait values. For example, we do not explicitly stipulate that $s_\chi(g)$ decreases as $\mu_\chi(g)$ increases. However, a mechanistic link between the mean and variance of a phenotypic trait could occur in more realistic situations where traits are constrained by trade-offs. For example, there is usually a fitness cost for a trait to deviate far from the mean, especially when the mean trait values are already optimized for a given environment.

In the abstract model, the effect of threshold is included by considering an upper bound on $\chi^{(i)}$ for each individual and on the mean phenotype μ_χ of each genotype, and the threshold values for both are set to be random numbers generated for each individual from $\mathcal{N}(\mu_{\chi_c}, \eta^2_{\chi_c})$. Due to the threshold, both the distribution of isogenic fluctuations and the distribution of genotype variations are set to be truncated normal distributions. As a result, the range over which the mean trait μ_χ and the trait χ can evolve becomes

smaller when μ_χ gets closer to the threshold, and therefore it automatically develops an effective 'slowing down' of the rate of evolution. The reason that we do not focus on lower thresholds of phenotype is because it does not matter for directed evolution that evolves in the direction of larger phenotype. If the directed evolution is designed to evolve in the opposite direction then it is the lower bound of phenotype that determines the evolution of phenotypic fluctuations. The model can also be extended in principle to the case where the selected trait depends on two or more phenotypes, and the overall threshold would be determined by the combination of higher or lower thresholds of each trait. We assumed that the timescale for changes in the threshold μ_{χ_c} is very long and set this to be a constant in all simulations. Therefore the value of the phenotype χ for a particular genotype g is distributed with a truncated normal distribution with an upper bound which is approximately $\mu_{\chi_c} \pm \eta^2_{\chi_c}$. We set $\mu_{\chi_c} = 100$ and $\eta_{\mu_\chi} = 3$.

4.4. Results of simulations

Figures 6(A) and (B) shows the evolution of the distribution of χ in the amplified population of the N_l individuals after each selection round, denoted by $P_p(\chi)$, under strong selection (panel (A)) and weak selection (panel (B)). The phenotypic fluctuation (or the cell-to-cell variation), given by σ_χ, is defined as the standard deviation of $P_p(\chi)$, and the average phenotype, represented by $\overline{\chi}$, is the expectation value of this distribution. In the remainder of this section, we will walk through the results of the simulations in detail, because there are a number of distinct cases that need to be presented. In the following section, we will interpret the outcome in terms of the behavior of the isogenic phenotype distributions of genotypes.

The evolution of σ_χ and $\overline{\chi}$ are shown in figures 6(C)–(E). Before round 40, the simulation results are broadly consistent with our naïve prediction, where strong and weak selection leads respectively to increase and decrease in σ_χ. Specifically, figure 6(A) shows an example of strong selection, where the selection strength is the strongest, defined here to be the case that the individuals with the top N_s largest phenotypes are selected ($N_r = N_s \ll N_f$). In this case $P_p(\chi)$ quickly evolves to large χ and becomes wider before round 5, and $\overline{\chi}$ and σ_χ increase accordingly as shown by the green curves in figures 6(D) and (E) which are averaged over 20 realizations. After $P_p(\chi)$ approaches μ_{χ_c} around round 5, $P_p(\chi)$ evolves slower and becomes left-skewed and slightly narrower, but still remains wider than the founder distribution, indicating saturating $\overline{\chi}$ and a slight decrease in σ_χ which is still larger than the variance in the founder strain as shown in figures 6(D) and (E). However, the increase in the population variance σ_χ is because of selection of large s_χ (reflected by increasing $\bar{s}_\chi$ in the green and orange curves in figure S3(B)) near the threshold, instead of due to the selection of both large and small s_χ as would be predicted by the naïve argument in section 4.2. We note that even though we do not assign any specific functional form for the asymmetry, but only assume truncation to Gaussian isogenic fluctuation distributions, the randomness in threshold automatically leads to smooth and skewed distributions like those we observed in the experimental data.

On the other hand, the case under weak selection, where $N_r/N_f = 0.5$ in figure 6(B), shows a different evolutionary trend from the case under strong selection, as would be predicted by the naïve argument. Figure 6(B) shows an example where $N_r/N_f = 0.5$. In the simulation, it takes longer (about 15 rounds) for $P_p(\chi)$ to approach μ_{χ_c} and to increase its width. Similarly, $\overline{\chi}$ saturates more slowly, and σ_χ only evolves to slightly larger values, as shown by the blue curves in figure 6(D) and (E) respectively. The increase in σ_χ is due to a different reason than in the case under strong selection as above: here both genotypes with large and small s_χ can be selected because selection is weak, leading to almost unchanged $\bar{s}_\chi$ (the blue curve in figure S3(B)).

In addition, in our numerical simulations we found a scenario which is not predicted by the naïve argument. If the selection strength is neither very strong nor weak, but has a model-dependent intermediate value, the variance initially increases but decreases later due to the accumulation of random mutations. This can lead to very different evolutionary trajectories from one simulation realization to another (orange curves after round 40 in figure 6(C)), and thus the population variance σ_χ can either increase or decrease depending in an unpredictable way on the selection round (orange curve in figure 6(E)).

The final average σ_χ evolving after 120 rounds is shown in figure 6(F) as a function of selection strength, with different sample population (N_s) and generation numbers during population amplification (m). The weaker the selection strength is, the smaller the final σ_χ becomes, because the probability of mutants with small s_χ being selected is higher. Similarly, larger sample population and more generations during population amplification allows more mutations with small s_χ to accumulate in the population and thus leads to smaller σ_χ. Except for the cases under very strong selection (e.g. $N_r \leqslant 2N_s$ or $N_r/N_s \leqslant 1/2^9$), the final σ_χ after many selection rounds declines compared with the standard deviation in χ of the founder strain which is represented by the red dashed line.

We also observed that if the traits are not bounded by a threshold, i.e. as $\mu_{\chi_c} \to \infty$, the traits evolve without bound in the simulations. Accordingly there is no saturation and there is no saturation of trait value after repeated rounds of selection, and there is no decline in the variance in the population. We note that besides selection strength the result can also depend on other parameters. For instance, if the mother-daughter correlation is high or the mutation range of the isogenic fluctuations (η_{s_χ}) is small, s_χ does not mutate enough

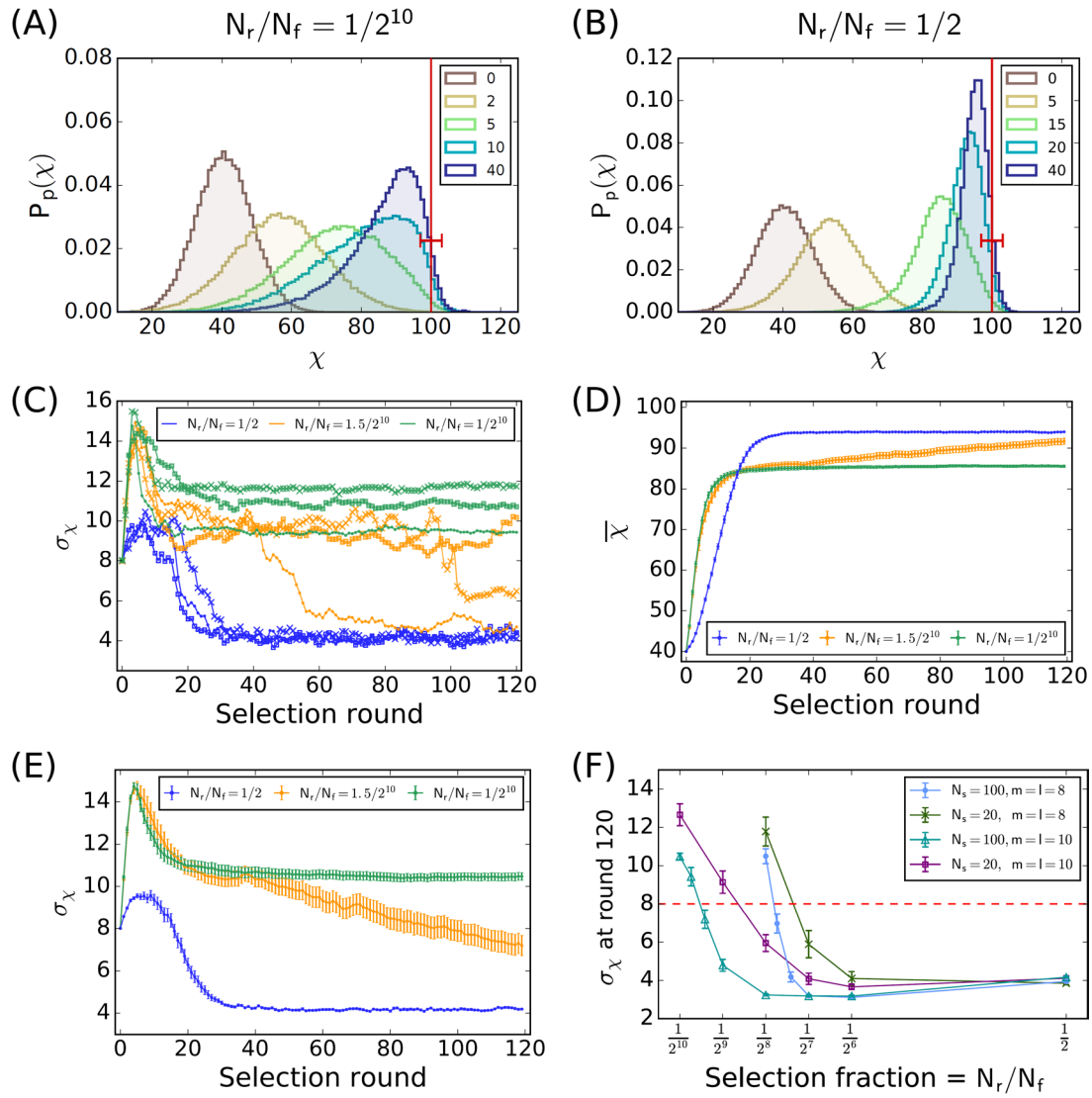

Figure 6. Simulations of the abstract model : (A) the distribution of χ of N_l individuals at different rounds under the strongest selection (e.g. the top N_s individuals are selected and $N_r/N_f = 1/2^{10}$): the distribution $P_p(\chi)$ quickly evolves and reaches μ_{χ_c}, and its width remains larger than the founder population even after $P_p(\chi)$ reaches the threshold around round 5, which implies that the genotypes with smaller s_χ are not particularly selected under strong selection. μ_{χ_c} and $2\eta_{\chi_c}$ are denoted by the vertical and horizontal red line. (B) $P_p(\chi)$ under very weak selection (e.g. $N_r/N_f = 1/2$): when $P_p(\chi)$ is away from μ_{χ_c}, its width increases when the mean of $P_p(\chi)$ evolves. The tail of $P_p(\chi)$ reaches μ_{χ_c} slower than the case of strong selection (around round 15), and after that $P_p(\chi)$ becomes tilted and narrower, indicating the overall variance in χ first increases and then declines eventually.
(C) The evolutionary trajectories in three simulation replicate (differentiated by different marker shapes) under different selection strength (blue: $N_r/N_f = 1/2$; orange: $N_r/N_f = 1.5/2^{10}$; green $N_r/N_f = 1/2^{10}$) show very stochastic dynamics due to random mutations. The overall mean and the standard deviation of χ are shown in (D) and (E) respectively, averaged over 20 replicates. If the selection strength is not extremely strong or weak, the increasing variance can decrease randomly due to accumulated mutations, which can lead to either increase or decrease in the final variance depending on the number of selection rounds (orange curve in (E)). (F) The overall variance in χ at round 120 is measured as a function of selection strength. The final variance is smaller for larger sample population (N_s) and more generations between selection rounds (m) due to the accumulation of more mutations. The red dashed line represents the standard deviation of χ for the founder strain. Parameters in the above simulations: $\mu_\chi(0) = 40, s_\chi(0) = 8, \mu_{\chi_c} = 100, \eta_{\mu_\chi} = 3, \eta_{s_\chi} = 1, \eta_{\chi_c} = 3$, and the stochastic values of χ, μ_χ and s_χ are binned with bin sizes equal to 1, 3 and 1 respectively. In (A)–(E) $N_s = 100$ and $m = l = 10$.

to increase much while μ_χ still evolves to the threshold, and therefore σ_χ can remain small even under very strong selection in this case.

In conclusion, through the simulations of this abstract model for directed evolution we have shown that an upper bound of phenotype can lead to finite-time saturation of the evolving phenotype, and to the decrease of cell-to-cell variation under temperate selection with typical parameter values. In the case with strong selection, the decrease of cell-to-cell variation is not a necessary consequence of the directed evolution procedure. Under strong selection, genotypes with large phenotypic fluctuations are favored, and the average phenotype and genotype values increase faster (figure 6(D)). In this sense, strong selection can be regarded as increasing the evolvability. In other words, whether phenotypic variation is advantageous or unfavorable depends on the selection strength and constraints on the phenotype.

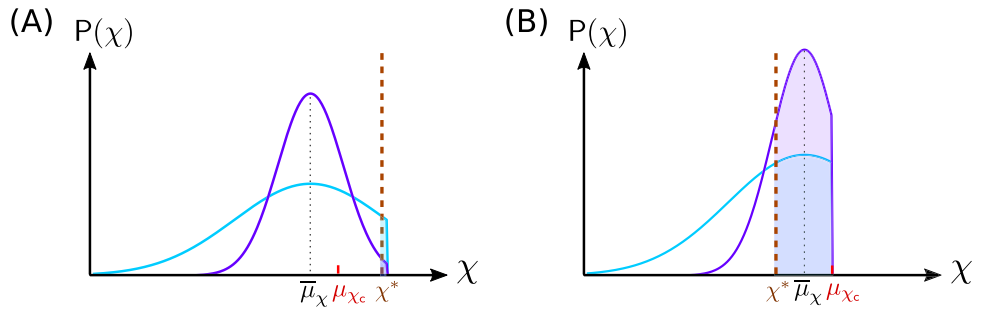

Figure 7. Heuristic prediction of the cell-to-cell variation in the abstract model. Illustration of selection when the phenotype distribution $P_p(\chi)$ approaches μ_{χ_c} under different strength of population bottleneck selection. (A) Under very strong selection where $N_r \sim N_s$, the genotypes with larger isogenic fluctuations s_χ and χ_c contribute larger phenotype χ at the tail of their distributions $P(\chi)$. Once $P_p(\chi)$ approaches μ_{χ_c}, $P(\chi)$ becomes truncated by the threshold, and the selction point χ^* is very close to the right end of the dominant genotypes (cyan curve). Before the next bottleneck selection, μ_χ of the most mutants is bounded by the random threshold with an average value of μ_{χ_c}, and the mutants with smaller s_χ (purple curve) have less density above the selection point χ^* compared with the dominant genotypes. At the next bottleneck selection, it is extremely unlikely for such a mutation to result in a phenotype in the small interval above χ^*, and thus the final variance remains large since strong selection favors those genotypes with substantial probability density above χ^*. (B) On the other hand, if selection is weak, where N_r/N_f is large ($\sim$0.5), genotypes with large s_χ or large χ_c are not particularly favored. However, mutations in μ_χ can develop heterogeneity in $P_p(\chi)$ and thus leads to an increasing variance. When $P_p(\chi)$ evolves near μ_{χ_c}, the distance between the selection point χ^* and the average truancation point μ_{χ_c} is large. Genotypes with large s_χ (cyan curve) no longer provide high density above the threshold for selection χ^*, but instead have less substantial probability of exhibiting phenotypes *below* this threshold (shaded cyan area). Further mutants with $\chi^* < \mu_\chi < \mu_{\chi_c}$ and smaller s_χ are favored under selection. The result is selection for genotypes with smaller s_χ, which leads to higher average χ and smaller final variance. When selection is not very strong or weak, genotypes with larger s_χ can be more favored at first, but after $P_p(\chi)$ approaches μ_{χ_c} the rare mutants with smaller s_χ can still be selected and have large probability density above the selection point even though they are unlikely to provide phenotypes at the right tail of $P_p(\chi)$. Therefore in this case, depending on how many mutants have accumulated, the final variance of $P_p(\chi)$ can either increase or decrease and should be a function of selection strength and the number of selection rounds.

4.5. Heuristic interpretation of the simulation results

Now that we have described the simulation results, we interpret them heuristically in terms of the isogenic phenotype distributions of genotypes. We emphasize that this is a *post hoc* rationalization of what the simulations revealed, and we cannot simply predict *a priori* these outcomes from naïve arguments. To understand the simulation results, we consider carefully the interplay between selection, mutation and random threshold. Here we refer to the lower bound on χ for the selected N_r individuals to be χ^*.

Without any physically-determined threshold on phenotype χ, the genotypes with larger isogenic fluctuations s_χ and χ_c can provide phenotypes with larger χ at the tail of their distributions $P(\chi)$. Therefore, under strong selection that acts at the right tail of distributions, the genotypes with larger isogenic fluctuations are more likely to be selected, and the variance of the distribution of phenotypes for the entire population $P_p(\chi)$ after population amplification will increase. On the other hand, if selection is weak, genotypes with large s_χ or large χ_c are not particularly favored, but mutations in μ_χ could develop heterogeneity in $P_p(\chi)$ and this leads to an increasing variance.

When the mean of $P_p(\chi)$ approaches μ_{χ_c}, $P(\chi)$ becomes truncated by the threshold, as illustrated in figure 7. Under strong selection, the genotypes with large s_χ and also large χ_c contribute the largest χ in $P_p(\chi)$, and μ_χ saturates quickly. When selection is extremely strong, e.g. $N_r \sim N_s$ and N_r/N_f is very small, the selection point χ^* is very close to the right tail of the dominant genotypes (cyan curve in figure 7 (A)). Before the next bottleneck selection, μ_χ of the most mutants is constrained by the random upper bound with an average value μ_{χ_c}, and therefore mutant genotypes with smaller s_χ (e.g. purple curve in figure 7(A)) have less density above χ^* compared with the dominant genotypes. At the next bottleneck selection, it is extremely unlikely for such a mutation to result in a phenotype in the small interval above χ^*, and thus the final variance remains large since strong selection favors those genotypes with substantial probability density above χ^*.

On the contrary, under weak selection where $N_r/N_f \sim$0.5, when $P_p(\chi)$ approaches μ_{χ_c}, the distance between the selection point χ^* and the average truancation point μ_{χ_c} is large (figure 7(B)). In this case, genotypes with large s_χ (cyan curve in figure 7(B)) no longer provide high density above the threshold for selection χ^* and instead have lower probability of exhibiting phenotypes *above* this threshold (shaded cyan area in figure 7(B)). Therefore selection favors mutants with $\chi^* < \mu_\chi < \mu_{\chi_c}$ and smaller s_χ. The result is selection for genotypes with smaller s_χ, which leads to higher average χ and smaller final variance. If the selection is not very strong or weak, genotypes with larger s_χ can be more favored at first, but after $P_p(\chi)$ approaches μ_{χ_c} the rare mutants with smaller s_χ can still be selected and have large probability density above the selection point even though they are unlikely to contribute phenotypes at the right tail of $P_p(\chi)$. Whether the final variance increases or decreases would depend on how many mutants have appeared and fixed. Since mutations are

rare and occur stochastically the final variance of $P_p(\chi)$ is expected to vary as a function of selection strength and the number of selection rounds.

A similar result that phenotypic variation could decrease (or increase) under weak (or strong) directed selection was found in [34] with a restricted bi-allele multi-loci model, for eight rounds, and the overall phenotype fluctuation of the population was assumed to be described by the mean and variance even after selection, and therefore could not capture the skewness effect. Also the effects of threshold and saturation of traits were not included [34, 35].

In general, a reduction in phenotypic fluctuations could be interpreted as stabilizing selection due to canalization [36], but the mechanism in this case is different from ours because there is no explicit threshold present. In the case of canalization, specific biological buffering mechanisms such as capacitance [37] are more likely to be at work. In short, our simulations suggest an alternative mechanism for phenotypic variation, arising as a generic consequence of bounded phenotypic variation under strong or weak selection.

4.6. Comparison between the experiment and the abstract model

The experimental results show that the variance of the run speed decreases with the number of rounds of selection, a result that our model predicts to occur when selection is weak. How can we estimate whether or not our experiment is truly in the weak selection regime? A naïve measure of the selection strength is the ratio N_r/N_f which we estimate to be order 10^2 in the experiment. Does this indicate strong selection then? It is difficult to draw a clear conclusion about this because, in general, selection acts on the phenotype space. The selection strength should be defined including the weighting of phenotype values, and not simply the number fraction that assumes equal weighting of each phenotype. In our experiment selection was applied in real space on agar plates, and thus the real physical phenotype that is being selected is a compound trait of multiple variables. Therefore, the selection fraction in the abstract model might not be simply related to the selection strength in the physical system. Thus, in order to test how the trend of phenotypic variance evolves with selection strength, it would be necessary to perform another set of experiments with different selection strengths, either a smaller selection fraction or selecting at different part of the population profile, to compare with the current experimental result shown in figure 4. In addition, the abstract model considers selection and evolution of a low-level trait of individuals instead of an emergent trait at population level. To explicitly compare with the experiment, we could extend our model by including two or more phenotypes and study the combined effect. For example, since the selection on colony is applied on the spatial position in our experiment, we may regard the selected property as dominance of length scale, which could be a function of run speed, tumble frequency and growth rate in the case of colony expansion. In the experiment, selection is applied on migration of the whole population, which is the property resulted from combined selection of individual chemotaxis and growth between two bottleneck selections. To explicitly include these features, it will require more variables and parameters, such as nutrient concentration and trait-dependent uptake rate which mimics the selection due to chemotaxis. These are planned for a future publication.

5. Biological mechanisms

Our abstract simulation makes a clear prediction about how phenotypic fluctuations should evolve in the presence of constraints on phenotypes under selection. Figure 2 shows that over the course of selection the swimming speed of the cell saturates at approximately 28 μm s^{-1} and does not change between rounds 10 and 15 of selection. This suggests the possibility that $|v_r|$ is in fact bounded from above in a manner similar to our evolutionary simulations. We note that the precise mechanism of this constraint is not known, but may be hydrodynamic, metabolic or genetic in origin. For example, the swimming speed increases with flagellar bundle rotation rate [38] which depends on the proton motive force and the pH, both of which depend on the metabolic state of the cell. Swimming speed is also under genetic regulation through a braking mechanism acting on the flagellar motors [39]. These mechanisms likely impart an upper bound on the swimming speed of the cell; indeed such a bound must exist given the finite propulsive force supplied by the flagella. Since we observe a saturation in swimming speed between rounds 10 and 15 of selection (figure 2(C)) and a concurrent decline in phenotypic fluctuations for $|v_r|$ (figure 4) we speculate that this reduction has as its basis a dynamic similar to our abstract model (figure 6), whereby the swimming speed is evolving towards an upper bound.

While swimming speed ($|v_r|$) appears to evolve towards an upper bound we observe a decline in run durations during selection as well as a decline in the phenotypic fluctuations in τ_r and τ_t (figures 2 and 4). It is less clear that explicit bounds apply to run and tumble durations. Indeed, mutants which exhibit very long or very short run durations have been isolated. Moreover, phenotypic fluctuations in the temporal statistics of runs and tumbles have been studied in *E. coli* for decades, and the molecular origins of these fluctuations are well understood. Since the seminal work of Koshland and Spudich [4], we now know that copy number fluctuations of the enzyme *cheR* and *cheB* drive large fluctuations in the run-tumble statistics at the single motor and single cell level [5, 26, 40]. Dufour *et al* [40] measured both gene expression and run-tumble statistics in single-cells to show a reduction in phenotypic fluctuations with increasing [*CheR*] and

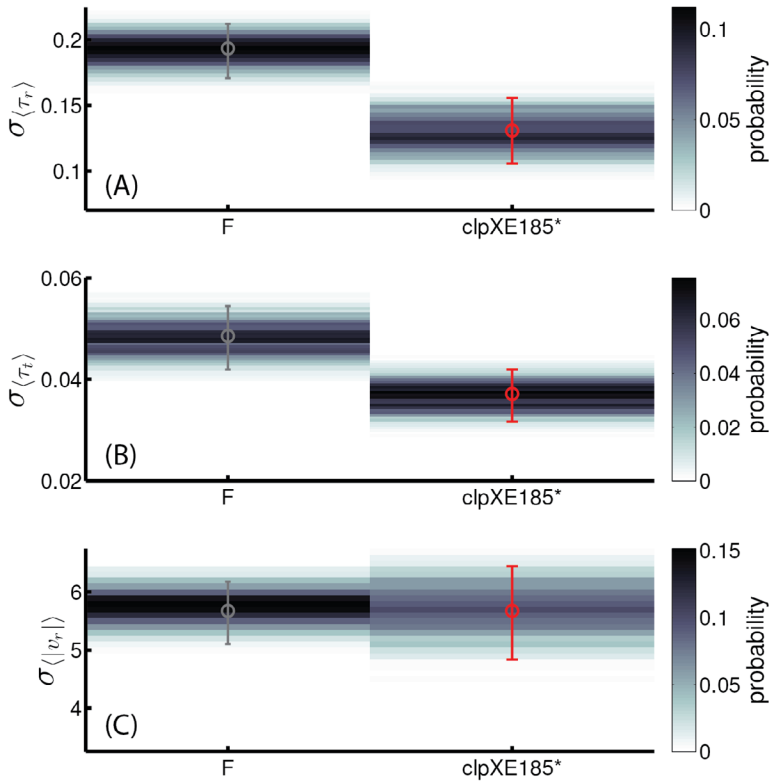

Figure 8. Cell-to-cell behavioral fluctuations in *clpX* mutant: individuality for a mutant with the *clpX*E185* mutation compared to the founder. Individuality for each population for (A) τ_r (B) τ_t and (C) $|v_r|$. We compute $\langle\tau_r\rangle$, $\langle\tau_t\rangle$ and $\langle|v_r|\rangle$ for each individual tracked and a standard deviation across individuals for each parameter (σ_χ). Data from 140 founder cells is reproduced from figure 4 and compared to 82 *clpX*E185* cells. Panels are identical to figure 4 with circles showing the sample σ_χ for each population. 95% confidence intervals from bootstrapping for each population are given by the error bars. Colormap shows the probability distribution of σ_χ from bootstrapping. Note distinct colorbars for each panel. The *clpX*E185* strain exhibits a statistically significant decline in $\sigma_{\langle\tau_r\rangle}$ and $\sigma_{\langle\tau_t\rangle}$ ($p < 0.01$, permutation test), but not $\sigma_{\langle|v_r|\rangle}$.

[*CheB*] concentrations *in vivo*. Phenotypic fluctuations declined when concentrations of both proteins increased while the ratio [*CheR*]/[*CheB*] remained constant [40]. Furthermore, increasing expression of both genes resulted in an increase in tumble frequency precisely as we observe in our selection experiment [40]. In a separate study, Vladimirov *et al* [41] show that the expression levels of both *CheR* and *CheB* are higher at the periphery of a colony expanding through 0.27% agar than at the center. Taken together these studies suggest that increasing *CheR* and *CheB* expression should reduce phenotypic fluctuations in τ_r and τ_t and that this reduction is correlated with distance from the center of the colony.

In light of this understanding we examined the mutations present in strains after selection. We performed whole genome sequencing on the ancestral strain as well as populations isolated after 5, 10 and 15 rounds of selection for four replicate selection experiments [22]. In every replicate we observe an identical mutation at >70% abundance by round 5 and fixed by round 10: a single nucleotide polymorphism which inserted a stop codon at position 185 in the 424 residue *ClpX* protein (*clpX*E185*). *ClpX* is the specificity subunit of the *ClpX-ClpP* serine protease, which degrades many target proteins including *FlhDC*. *flhDC* is the master regulator of a coherent feedforward motif which governs the expression of motility and chemotaxis genes including *cheR* and *cheB* enzymes, which are determinants of phenotypic fluctuations [42].

To investigate the role of the mutation we observed in *clpX* in phenotypic fluctuations, we reconstructed the *clpX*E185* mutation in the ancestral background using scarless recombineering. We confirmed that this mutation alone is sufficient to drive faster migration through increasing run speed and decreasing growth rate [22]. Moreover, this mutation alone causes a decrease in the phenotypic fluctuations in run duration and tumble duration, but not run speed relative to the ancestral population (figure 8).

We considered whether the mutation we observe in *clpX* might logically result in increased levels of *cheR* and *cheB* and therefore the reduced phenotypic fluctuations we observe. Previous studies have shown that mutations in *ClpX* increase levels of *FlhDC* in the cell [43]. Zhao *et al* [44] show that deleting *flhDC* results in substantial reduction in expression of the downstream *cheR/B* genes. However, inducing *FlhDC* expression above wild-type levels appears not to increase expression of downstream genes substantially [44]. Despite this, single cell measurements show a positive correlation between *flhC* and *cheY* expression levels [45].

Since *cheY* is co-transcribed with *cheR* and *cheB* we speculate that increases in *FlhDC* levels in the cell may drive increases in *cheR* and *cheB* expression and that could reduce phenotypic fluctuations. Further studies are needed to directly measure the *meche* operon expression levels in the presence and absence of the *clpX* mutation we observe. Since we cannot stipulate whether expression of the relevant genes is subject to a bound, we cannot conclude that the mechanism proposed in our abstract model describes the decline in fluctuations in run and tumble durations.

The *clpX*E185* mutation alone drives an increase in run speed to 24.2 μm s^{-1} from 18.2 μm s^{-1} for the founder whereas the average run speed of the round 15 evolved strain is 28.7 μm s^{-1} [22]. These results suggest that the mutant run speed is, on average, far from the apparent upper bound in swimming speed. As our abstract model would predict for the mutant, we observe no decrease in $\sigma_{\langle|v_r|\rangle}$ in the mutant relative to the founder—potentially because the mutant phenotype is not constrained by an upper bound on run speed.

6. Discussion

In our measurements we report that selection drives reduction in phenotypic fluctuations associated with chemotactic mobility. We also identified the mutations that appear to be implicated in this evolution of phenotype fluctuations. Are the results surprising, or could they have been predicted on general grounds related to the fluctuation-dissipation theorem and other global properties of stochastic gene expression?

Our abstract model suggests that such a reduction may arise from selection in the presence of a constraint on phenotypes. From our numerical simulations, we show that the phenotype variation in a minimal model of directed evolution evolves as a function of the selection strength of population bottleneck and the number of selection rounds. Within a broad range of parameter, the variance increases under extremely strong selection that always chooses the top N_s largest phenotypes near the threshold where N_s is the sample population at each selection round, while temperate selection allows accumulation of mutants with small isogenic fluctuations and hence can lead to decrease in the variance. Thus our data suggest the possibility that swimming speed may be constrained in *E. coli* by biophysical or metabolic means. Since there is no direct evidence for a threshold on traits such as run and tumble duration, the reduction in phenotypic fluctuations in run and tumble durations in our data could have a distinct mechanistic basis which may not be captured by our simple abstract model. Another possible explanation for the reduced variance in run and tumble duration is that these traits evolve to lower values and are bounded by some lower bound, since cells cannot have infinitesimal run and tumble duration due to physical limitations. This could make sense because in the soft agar gel the more frequently and the more quickly a cell switches its direction, the more efficiently it could find the correct gradient to do chemotaxis. This is consistent with the experimental results shown in figure 4(A), where $P_p(\langle\tau_r\rangle)$ becomes more right-skewed as the mean of $P_p(\langle\tau_r\rangle)$ decreases over time, which is similar for $\langle\tau_t\rangle$. Nevertheless, in figure 4(A) the left tail of $P_p(\langle\tau_r\rangle)$ does not clearly evolve towards the left even though $\overline{\langle\tau_r\rangle}$ decreases. We suspect that the main phenotype subjected to the threshold could be a composite trait such as the run length which is the multiplication of run speed and run duration, and therefore the evolutionary trajectory of a single trait could become non-monotonous over time. Further work is needed to elucidate the role of constraints on phenotypic fluctuations in run and tumble duration.

In addition, even though experimental data show a small increase or no significant decrease in variance between rounds 10 and 15 (figures 4(D)–(F)), the variance at round 15 is always less than variance of the founder and the increase is only significant for $\langle\tau_t\rangle$. Therefore, this increase in variance from round 10–15 is at the limits of detectability and statistical significance in our experiments. In our study of the abstract model as presented in section 4.4, when the mother-daughter correlation is high or the mutation range of the isogenic fluctuations (η_{s_χ}) is small, phenotype variance can decrease even under very strong selection. In these cases, if mutation rate is high or the fluctuation range in threshold (η_{χ_c}) is large, strong selection can still select mutant genotypes with large isogenic fluctuations (s_χ) once mutants accumulate enough when $\overline{\chi}$ has evolved to near χ_c, which can cause the variance in $P_p(\chi)$ to 'bounce back' and increase again as shown in figure S3. If the model includes an intrinsic tendency to decline with increasing phenotype mean and the population amplification step (growth for m generations) is not long enough to eliminate bias in the phenotype due to mother-daughter correlation, the bounce-back in variance could also appear due to the selection increasing the mean to a point whereby the intrinsic phenotype variance is smaller than the population phenotype variations. Another logically-allowed possibility for the increase in variance is that the mutants that begin to dominate in the population at later rounds of selection have larger variance than the ones at earlier stages. Finally, the specific constraint on the distributions due to the upper bound can also change the final variance. However, these possibilities are parameter-dependent and thus are not necessary at the current stage especially since it is uncertain that the bounce-back is robust in the experiment.

Our abstract model of directed evolution applies to a broad range of potential systems and makes predictions of possible scenarios as to how the strength

of selection can influence phenotypic fluctuations. Genetic, biophysical and chemical constraints play an important role in the dynamics of biological systems from higher organisms such as fungi [46] to limits on the speed of protein translation [47] and enzyme specificity [48]. Our study highlights the potentially important role for these constraints in determining the limits of phenotypic fluctuations. Future experimental evolution work could exploit known phenotypic constraints and directed evolution to directly test the predictions of our model.

At a lower level of biological organization the mechanisms underlying phenotypic fluctuations remain hard to uncover in general due to the complex relationship between gene expression, protein function and cell-level phenotypes. Despite the difficulty of connecting phenotypes to gene expression recent work has shown universal statistical properties in protein copy number distributions, with monotonically increasing scaling of the variance in protein abundances with mean expression levels [13–15, 49]. These universal properties of protein abundance fluctuations may provide a basis for understanding the evolution of phenotypic fluctuations in situations where the relevant regulatory architecture is known [9, 10]. However, at present, a molecular accounting for the mechanism of the evolution of phenotype fluctuations requires detailed knowledge of the signaling pathways at work. Our hope is that in studying abstract models such as the one presented here, we may uncover a more general understanding of when and why phenotypic fluctuations evolve.

Acknowledgments

We thank Alvaro Hernandez at the Roy J Carver Biotechnology Center at the University of Illinois at Urbana-Champaign who performed the HiSeq sequencing and Elizabeth Ujhelyi who provided assistance with MiSeq sequencing. We acknowledge partial support from National Science Foundation through grants PHY 0822613 and PHY 1430124 administered by the Center for Physics of Living Cells. H-Y Shih was partially supported by the NASA Astrobiology Institute award NNA13AA91A, the L S Edelheit Family Fellowship in Biological Physics (2016) and the Government Scholarship for Study Abroad (2015-2016), Taiwan.

ORCID iDs

Hong-Yan Shih https://orcid.org/0000-0002-3578-7785
David T Fraebel https://orcid.org/0000-0002-8668-3476
Nigel Goldenfeld https://orcid.org/0000-0002-6322-0903
Seppe Kuehn https://orcid.org/0000-0002-4130-6845

References

[1] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6
[2] Lehner B 2010 Conflict between noise and plasticity in yeast *PLoS Genet.* **6** e1001185
[3] Balaban N Q, Merrin J, Chait R, Kowalik L and Leibler S 2004 Bacterial persistence as a phenotypic switch *Science* **305** 1622–5
[4] Spudich J and Koshland D 1976 Non-genetic individuality: chance in the single cell *Nature* **262** 467–71
[5] Frankel N W, Pontius W, Dufour Y S, Long J, Hernandez-Nunez L and Emonet T 2014 Adaptability of non-genetic diversity in bacterial chemotaxis *eLife* **3** e03526
[6] Baldwin J M 1896 A new factor in evolution *Am. Nat.* **30** 441–51
[7] Waddington C H 1953 Genetic assimilation of an acquired character *Evolution* **7** 118–26
[8] West-Eberhard M 1989 Phenotypic plasticity and the origins of diversity *Ann. Rev. Ecol. Evol. Syst.* **20** 249–78
[9] Sato K, Ito Y, Yomo T and Kaneko K 2003 On the relation between fluctuation and response in biological systems *Proc. Natl Acad. Sci. USA* **100** 14086–90
[10] Ito Y, Toyota H, Kaneko K and Yomo T 2009 How selection affects phenotypic fluctuation *Mol. Syst. Biol.* **5** 264
[11] Yoshida M, Tsuru S, Hirata N, Seno S, Matsuda H, Ying B-W and Yomo T 2014 Directed evolution of cell size in *Escherichia coli BMC Evol. Biol.* **14** 257
[12] Furusawa C, Suzuki T, Kashiwagi A, Yomo T and Kaneko K 2005 Ubiquity of log-normal distributions in intra-cellular reaction dynamics *Biophysics* **1** 25–31
[13] Salman H, Brenner N, Tung C-k, Elyahu N, Stolovicki E, Moore L, Libchaber A and Braun E 2012 Universal protein fluctuations in populations of microorganisms *Phys. Rev. Lett.* **108** 238105
[14] Sherman M S, Lorenz K, Lanier M H and Cohen B A 2015 Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression *Cell Syst.* **1** 315–25
[15] Taniguchi Y, Choi P J, Li G-W, Chen H, Babu M, Hearn J, Emili A and Xie X S 2010 Quantifying *e. coli* proteome and transcriptome with single-molecule sensitivity in single cells *Science* **329** 533–8
[16] Brenner N, Braun E, Yoney A, Susman L, Rotella J and Salman H 2015 Single-cell protein dynamics reproduce universal fluctuations in cell populations *Eur. Phys. J.* E **38** 102
[17] Susman L, Kohram M, Vashistha H, Nechleba J T, Salman H and Brenner N 2016 Statistical properties and dynamics of phenotype components in individual bacteria (arXiv:1609.05513)
[18] Milo R, Hou J H, Springer M, Brenner M P and Kirschner M W 2007 The relationship between evolutionary and physiological variation in hemoglobin *Proc. Natl Acad. Sci. USA* **104** 16998–7003
[19] Barrick J E, Yu D S, Yoon S H, Jeong H, Oh T K, Schneider D, Lenski R E and Kim J F 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli Nature* **461** 1243–7
[20] Rainey P B and Travisano M 1998 Adaptive radiation in a heterogeneous environment *Nature* **394** 69–72
[21] Bachmann H, Fischlechner M, Rabbers I, Barfa N, Branco dos Santos F, Molenaar D and Teusink B 2013 Availability of public goods shapes the evolution of competing metabolic strategies *Proc. Natl Acad. Sci. USA* **110** 14302–7
[22] Fraebel D T, Mickalide H, Schnitkey D, Merritt J, Kuhlman T E and Kuehn S 2017 Environment determines evolutionary trajectory in a constrained phenotypic space *eLife* **6** e24669
[23] Wolfe A J and Berg H C 1989 Migration of bacteria in semisolid agar *Proc. Natl Acad. Sci. USA* **86** 6973–7
[24] Adler J 1966 Chemotaxis in bacteria *Science* **153** 708–16
[25] Yi X and Dean A M 2016 Phenotypic plasticity as an adaptation to a functional trade-off *eLife* **5** e19307
[26] Korobkova E, Emonet T, Vilar J M, Shimizu T and Cluzel P 2004 From molecular noise to behavioural variability in a single bacterium *Nature* **428** 574–8

[27] Bai F, Che Y-S, Kami-ike N, Ma Q, Minamino T, Sowa Y and Namba K 2013 Populational heterogeneity versus temporal fluctuation in *Escherichia coli* flagellar motor switching *Biophys. J.* **105** 2123–9
[28] Jordan D, Kuehn S, Katifori E and Leibler S 2013 Behavioral diversity in microbes and low-dimensional phenotypic spaces *Proc. Natl Acad. Sci. USA* **110** 14018–23
[29] Wagner G P and Altenberg L 1996 Perspective: complex adaptations and the evolution of evolvability *Evolution* **50** 967–76
[30] Lande R 1976 Natural selection and random genetic drift in phenotypic evolution *Evolution* **30** 314–34
[31] Jensen J L, Lake L W, Corbett P W M and Goggin D J 2000 *Statistics for Petroleum Engineers and Geoscientists* (Amsterdam: Elsevier)
[32] Raser J M and O'shea E K 2005 Noise in gene expression: origins, consequences, and control *Science* **309** 2010–3
[33] Feinberg A P and Irizarry R A 2010 Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease *Proc. Natl Acad. Sci.* **107** 1757–64
[34] Hill W G and Zhang X-S 2004 Effects on phenotypic variability of directional selection arising through genetic differences in residual variability *Genet. Res.* **83** 121–32
[35] Eldar A and Elowitz M B 2010 Functional roles for noise in genetic circuits *Nature* **467** 167–73
[36] Waddington C H 1942 Canalization of development and the inheritance of acquired characters *Nature* **150** 563–5
[37] Rutherford S L and Lindquist S 1998 Hsp90 as a capacitor for morphological evolution *Nature* **396** 336–42
[38] Darnton N C, Turner L, Rojevsky S and Berg H C 2007 On torque and tumbling in swimming *Escherichia coli J. Bacteriol.* **189** 1756–64
[39] Boehm A, Kaiser M, Li H, Spangler C, Kasper C A, Ackermann M, Kaever V, Sourjik V, Roth V and Jenal U 2010 Second messenger-mediated adjustment of bacterial swimming velocity *Cell* **141** 107–16
[40] Dufour Y S, Gillet S, Frankel N W, Weibel D B and Emonet T 2016 Direct correlation between motile behavior and protein abundance in single cells *PLoS Comput. Biol.* **12** e1005041
[41] Vladimirov N, Løvdok L, Lebiedz D and Sourjik V 2008 Dependence of bacterial chemotaxis on gradient shape and adaptation rate *PLoS Comput. Biol.* **4** e1000242
[42] Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette M and Alon U 2001 Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria *Science* **292** 2080–3
[43] Tomoyasu T, Takaya A, Isogai E and Yamamoto T 2003 Turnover of FlhD and FlhC, master regulator proteins for Salmonella flagellum biogenesis, by the ATP-dependent ClpXP protease *Mol. Microbiol.* **48** 443–52
[44] Zhao K, Liu M and Burgess R R 2007 Adaptation in bacterial flagellar and motility systems: from regulon members to 'foraging'-like behavior in *E. coli Nucl. Acids Res.* **35** 4441–52
[45] Løvdok L E 2008 Gene expression noise and robustness in the *Escherichia coli* chemotaxis pathway *PhD Thesis* University of Heidelberg (https://doi.org/10.11588/heidok.00009109)
[46] Roper M, Pepper R E, Brenner M P and Pringle A 2008 Explosively launched spores of ascomycete fungi have drag-minimizing shapes *Proc. Natl Acad. Sci. USA* **105** 20583–8
[47] Scott M, Gunderson C W, Mateescu E M, Zhang Z and Hwa T 2010 Interdependence of cell growth and gene expression: origins and consequences *Science* **330** 1099–102
[48] Savir Y, Noor E, Milo R and Tlusty T 2010 Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape *Proc. Natl Acad. Sci. USA* **107** 3475–80
[49] Brenner N, Newman C M, Osmanović D, Rabin Y, Salman H and Stein D L 2015 Universal protein distributions in a model of cell growth and division *Phys. Rev.* E **92** 042713

Supplementary Material for: Biophysical constraints determine the selection of phenotypic fluctuations during directed evolution

Hong-Yan Shih, Harry Mickalide, David T. Fraebel, Nigel Goldenfeld*, Seppe Kuehn*

1. Correlation between mother and daughter in the bacterial chemotaxis experiment

The correlation between mother and daughter could be effective to accumulate phenotypic variation during directed evolution if the correlation is high. The phenotype variance could decrease or increase less due to high mother-daughter correlation, which will be discussed in the next section. However, it was shown in the relaxation experiment in Fig. 1-figure supplement 5 in [1] that after about 140 generations in liquid environment where selection is absent that no additional mutations occurred and that the fast migration phenotype was not lost. For these results to be explained by mother-daughter correlation requires extremely high correlation (> 0.99) between mother and daughter. However, direct measurements of the correlation between mother and daughter cells show correlation coefficients around 0.5 (Fig. S1).

2. Comparison of the time scale in the experiment and the abstract model

In general, the time it takes for trait to saturate depends on many parameters such as mutation rate, selection strength, population size, generation number in each selection round, threshold value, growth rate and nutrient concentration etc. Since we want to explore the generic qualitative phenomena in directed evolution, the quantitative comparison is not the focus in our model. Moreover, to simulate with the same order of population size in the experiment which is $10^6 - 10^9$ would be too time-consuming and not practical for the purpose of the abstract model. For example, in simulation we set the mutation rate is much higher than the mutation rate in the experiment in order to observe mutations in a much smaller population within few selection rounds, and we set the threshold to be far away from the initial condition in order to see the effect of the threshold on the dynamics. Below we show an example of the simulation of the abstract model with similar saturation round number as the experiment.

We are aware of the simplification in the abstract model and have considered how minimal it could be to suffice as a qualitative demonstration for directed evolution. For

example, In our model we assume that the bottleneck selection only applies at the end of each round for simplicity, and the probability to replicate is one for all individuals. In other words, this version of the abstract model effectively assumes that the selection due to chemotaxis between two bottleneck selections is weak and shows that bottleneck selection is sufficient to describe evolution of phenotypic fluctuations in a minimal model.

Although the effect of finite nutrients and chemotaxis within each round before the bottleneck section is not included in the model, we have considered including a trait-dependent growth rate to differentiate between individuals with different trait values. If the individual growth rate is only determined by the absolute trait value (χ), the strength of selection attributed to growth rate could be either weak or strong, depending on how fast the growth rate saturates due to mutations. Another possibility is that the growth rate is determined by the relative trait value in the current population, which is more similar to the chemotaxis experiment. However, in the abstract model, selection is applied in the trait space instead of the real space, and therefore arbitrary mutations could lead to sudden jumps to high trait values in the trait space. Moreover, since mutations might occur to only few individuals, those mutants with overwhelming trait values will have a saturated growth rate, leaving most individuals with relatively small trait values in the population to have much smaller growth rate, and thus the whole population would quickly shrink. To fix the growth problem, it will require more variables and parameters, such as nutrient concentration and trait-dependent uptake rate which mimics the selection due to chemotaxis between two bottlenecks. Another easier but less straightforward modification could be including exclusive competition between individuals with different trait values. We conclude that including the role of growth rate in the model is a non-trivial extension of what we have already done.

Nevertheless, the goal of the abstract model is to understand the possible general causes with minimal factors that would affect the directed evolution of phenotypic fluctuations. In this sense it is reasonable to test if such a minimal model is sufficient to show different evolutionary trends. As we have proposed in the discussion section, further modifications to the model and the experiment could be studied in the future, which would not be possible without testing the simplest abstract model. In the future model, each modification still needs to be tested one by one to understand their role in evolution.

Under these approximations, the time scale in the simulation of our model depends mainly on replication rate, mutation rate and the effect size of each mutation. We set the probability of replication in each generation to be constant (which is one) in order to reduce the number of parameters.

3. Discussion of bounce-back in the variance

The increase in variance at round 15 is not statistically significant at least in run speed and run duration. Results of statistical tests demonstrating this fact have been added to the caption to Fig. 4. Also we do not see obvious bimodality in any of the phenotypic

trait distributions in Fig. 4(A-C). For run speed the increase in variance at round 15 is still inside the error of the variance at round 10. However the variance at round 15 is always less than variance of the founder. In short, we think that the increase in variance from round 10 to 15 is at the limits of detectability in our experiments, but we agree with the referee that it is interesting to try and understand if there is a natural mechanism for such a "bounce-back" effect. There are three possibilities that we had considered but did not feel confident enough in them to include them in our original manuscript.

The first possibility for the decrease and the bounce-back in phenotype variance is high mother-daughter correlation or small mutation range in isogenic fluctuations (η_{s_χ}) under strong selection ($N_r \gg N_s$ and $N_r/N_f \ll 1$). In these cases s_χ does not mutate enough to increase much, while μ_χ still evolves to the threshold, and therefore σ_χ can remain small even under strong selection. If mutation rate (ν) is high or the fluctuation range in threshold ((η_{χ_c})) is also large, once the phenotype values has evolved to near μ_χ, mutations will accumulate changes in genotype, and genotypes with large isogenic fluctuations could be favored under strong selection, and therefore the variance can increases again. However, this requires high enough mother-daughter correlation to make the variance decrease enough before it increases due to mutations. If the mother-daughter correlation or selection strength becomes weaker, it requires larger fluctuations to show the bounce-back in the later rounds, such as higher mutation rate or larger fluctuation range of threshold. Examples of simulations under such conditions are shown in Fig. S3 as below.

Moreover, at high mother-daughter correlation, if the generation number within each selection round, m, is small (e.g. $m > 4$), once rare mutations occur, there would not be enough time for them to build up large enough population in the end of the selection round, and therefore the population profile could develop into a bimodal shape with time, making the final variance increase further. However, we consider this effect as an artifact from limited generation number and population size, because in the experiment we did not observe bimodal distribution in traits, and both the generation number (which is about 10) and population size are not small.

In the experiment, the mother-daughter correlation is between 0.23 and 0.65, which is not particularly high. Since in the experiment it is not clear if the bounce-back is a generic effect and it is not easy to estimate the value of ν, η_{s_χ} and η_{χ_c}, we do not tend to use the above argument to claim that the bounce-back at round 15 in the current data was due to strong selection. Further measurements of more accurate mother-daughter correlation and with different strength of selection would be needed to test the model .

The second possibility is that the phenotype variance has an intrinsic tendency to decline with increasing phenotype mean. When we had included this in simulations, we could observe a bounce-back for some model parameters. This bounce-back is basically due to the selection increasing the mean to a point whereby the intrinsic phenotype variance is smaller than the population phenotype variations. The bounce-back can be ameliorated if the selected bacteria are allowed to increase their biomass in liquid culture

for enough generations, but in principle a residual bounce-back could still remain. This explanation is quite complicated and definitely parameter-dependent. For these reasons, we did not attempt a detailed characterization and exploration of this effect, in addition to the fact that the empirical results are at the limit of detectability and statistical significance.

Finally, another logically-allowed possibility for the increase in variance is that the mutants that begin to dominate in the population at later rounds of selection have larger variance than the ones at earlier stages. In the experiment , the *clpX* mutation is at or near fixation by round 5 in all replicates of the selection process. Therefore, to experimentally test this possibility would require that we study the phenotypic effects of mutations present in the population at later rounds of selection.

Given the high degree of uncertainty as to which, if any, of these possibilities is correct, we would prefer not to speculate in the manuscript, especially since it is uncertain that the bounce-back is robust in the experiment.

4. Evolution of genotype in the abstract model

We have checked that μ_χ always evolves and saturates to a value near reaches μ_{χ_c}, and therefore the increasing $\overline{\chi}$ is resulted from mutation but not due to extremely strong mother-daughter correlation (e.g. $\rho > 0.9$). The saturation value of $\overline{\mu_\chi}$ and the saturation speed depend on the selection strength. When χ is away from the mean threshold μ_{χ_c}, under the strongest selection strength (green curve in Fig. S4(A)) individuals with larger μ_χ are selected, and thus $\overline{\mu_\chi}$ increases faster than the weak selection strength case (blue curve in Fig. S4(A)). This is similar as the evolution speed of $\overline{\chi}$ in Fig. 6(D) in the manuscript.

However, as $\overline{\chi}$ approaches μ_{χ_c}, $\overline{\mu_\chi}$ also depends on the threshold χ_c which evolves accordingly with selection strength. Under the strongest selection, before $\overline{\chi}$ approaches μ_{χ_c}, individuals with large s_χ have already been selected, and once μ_χ evolves near μ_{χ_c}, further mutations, either with larger μ_χ or different s_χ are rare and can hardly contribute enough individuals to the far right tail of the population $P_p(\chi)$ to get selected. Therefore $\overline{\mu_\chi}$ saturates to a value with about an amount of the final σ_χ below μ_{χ_c}, similar as $\overline{\chi}$ under extremely strong selection. The evolution of $\overline{\mu_\chi}$ under weak selection is also similar as the evolution of $\overline{\chi}$.

On the other hand, if selection is not extremely strong (e.g. $N_r > N_s$), there is a finite possibility to select the rare mutants with larger μ_χ but not necessary with larger χ. This can allow $\overline{\mu_\chi}$ to evolve above μ_{χ_c} (orange curve in Fig. S4(A)). Nevertheless, the measurable and selectable observables are phenotypes χ, $\overline{\chi}$ and σ_χ, and we do not observe that $\overline{\chi} \sim \mu_{\chi_c}$ even though $\overline{\mu_\chi}$ becomes near or above μ_{χ_c}. This could be understood as following. Since in the abstract model the only selection comes from the population bottleneck at the end of each selection round, if selection is not extremely strong and the mother-daughter correlation is not very high, after many generations before the next selection, χ could relax back to smaller values (orange curve in Fig.

6(D) in the manuscript) even though μ_{χ_c} is large.

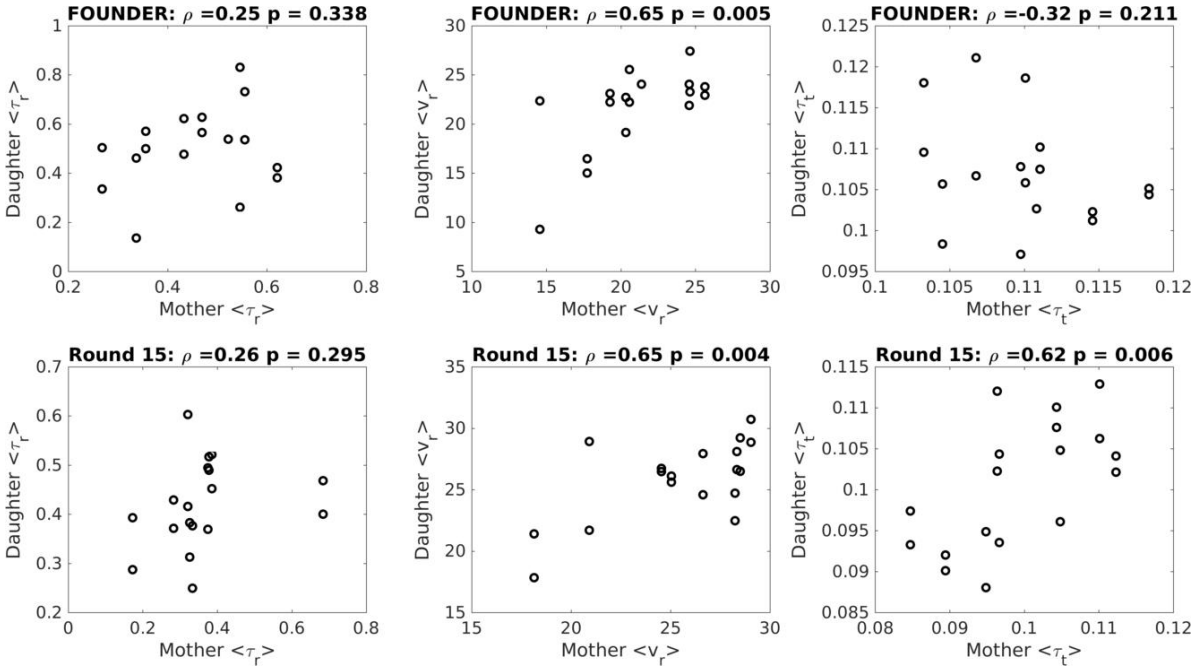

Figure S1. Mother-daughter correlation in the bacterial chemotaxis experiment: Mother-daughter correlations at the single-cell level for founding strain (top row) and round 15 evolved strain (bottom row). These data were obtained using the single-cell tracking approach described in the main text and in [1]. Single cells were loaded into microfluidic devices allowed to swim until division. The two resulting daughter cells were then tracked for the entirety of their lifetime. Average run duration, speed and tumble durations were computed for each individual. Data are presented for 17 mother-daughter pairs (founder strain) and 18 pairs (round 15 strain). Title of each panel reports the correlation coefficient and the associated p-value which is computed assuming the data arise from a bivariate normal distribution

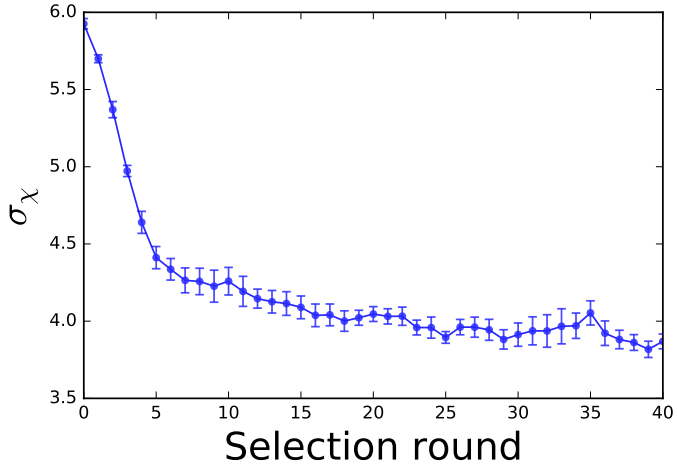

Figure S2. Abstract model simulation with comparable time scale as the experimental data: The initial values and the threshold value are set to be similar as the distribution of run speed in the data. Parameters: $N_s = 100, N_r/N_f = 1/2^3, m = l = 8, \rho = 0.8, \mu_\chi(0) = 20, s_\chi(0) = 6, \mu_{\chi_c} = 40, \eta_{\mu_\chi} = 3, \eta_{s_\chi} = 0.5, \eta_{\chi_c} = 3.5, \nu = 0.5$, and the bin sizes for χ, μ_χ and s_χ are 1, 1 and 0.5.

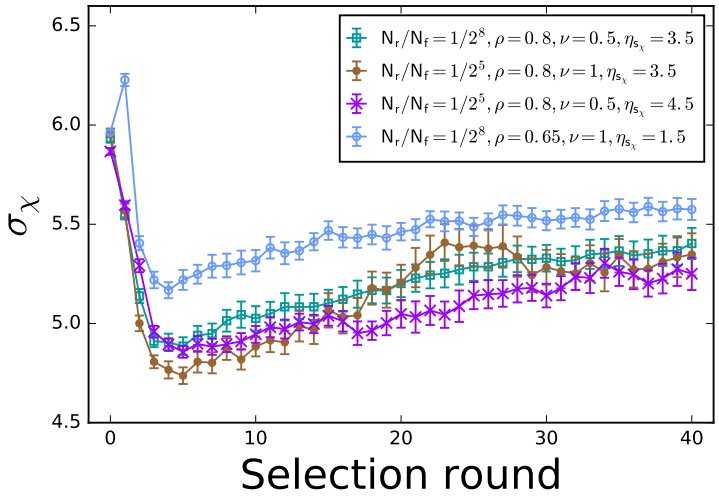

Figure S3. Abstract model simulation showing "bounce-back" in phenotypic variance: Abstract model simulations showing possible bounce-back in variance, with strong selection strength ($N_r \gg N_s$ and $N_r/N_f \ll 1$) and high mother-daughter correlation (ρ), large fluctuation range in threshold (η_{χ_c}) or high mutation rate (ν). Parameters: $N_s = 100, m = l = 8, \mu_\chi(0) = 20, s_\chi(0) = 6, \mu_{\chi_c} = 40, \eta_{\mu_\chi} = 3, \eta_{s_\chi} = 0.5$, and the bin sizes for χ, μ_χ and s_χ are 1, 3 and 0.5.

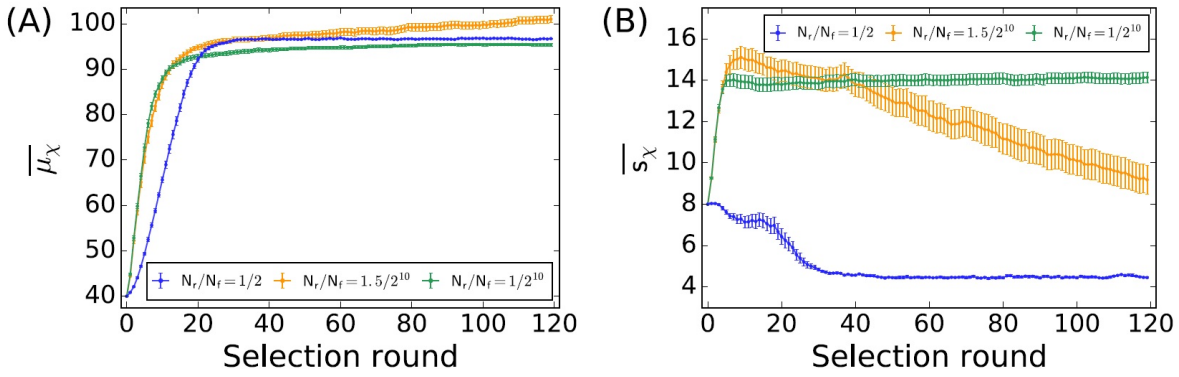

Figure S4. Evolution of genotype in the abstract model: The population average of (A) μ_χ and (B) s_χ in the abstract model simulations, from the same data shown in Fig. 6(D) and 6(E) in the manuscript. Before $\overline{\chi}$ reaches the threshold, if selection is weak, genotypes with large s_χ are not particularly favored, and thus $\overline{s_\chi}$ remains similar as the initial value.

[1] D. T. Fraebel, H. Mickalide, D. Schnitkey, J. Merritt, T. E. Kuhlman, and S. Kuehn, "Environment determines evolutionary trajectory in a constrained phenotypic space," *eLife*, vol. 6, p. e24669, Mar. 2017.