# Molecular Evolution under Fitness Fluctuations

Ville Mustonen and Michael Lässig

*Institut für Theoretische Physik, Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany*

(Received 22 November 2006; revised manuscript received 20 November 2007; published 10 March 2008)

Molecular evolution is a stochastic process governed by fitness, mutations, and reproductive fluctuations in a population. Here, we study evolution where fitness itself is stochastic, with random switches in the direction of selection at individual genomic loci. As the correlation time of these fluctuations becomes larger than the diffusion time of mutations within the population, fitness changes from an *annealed* to a *quenched* random variable. We show that the rate of evolution has its maximum in the crossover regime, where both time scales are comparable. Adaptive evolution emerges in the quenched fitness regime (evidence for such fitness fluctuations has recently been found in genomic data). The joint statistical theory of reproductive and fitness fluctuations establishes a conceptual connection between evolutionary genetics and statistical physics of disordered systems.

DOI: 10.1103/PhysRevLett.100.108101 PACS numbers: 87.23.Kg, 61.43.−j, 87.15.Cc

Biological evolution has produced a bewildering diversity of life. New species and functional innovations within a species arise as adaptive response to an ever-changing environment. It has become a major challenge for molecular evolutionary biology to trace the genomic basis of these changes [1]. How much of the DNA sequence divergence observed between different species encodes for adaptive changes, and how much is due to random mutations accumulated over time? In our genome, where are the specific changes underlying, e.g., the evolution of language and brain development, which make us human? And where does selection continue to act today? Obviously, these questions bear rich scientific implications, which range from the conceptual foundations of evolutionary biology to biomedical applications, e.g., the identification of genes associated with genetic diseases.

Despite a growing amount of genomic data available, identifying evolutionary adaptations is a difficult signal-to-noise problem, since the underlying sequence changes emerge from a complex stochastic process at the level of populations [2,3]. In the simplest case of a single locus (e.g., a base pair) with two states (alleles) $A$ and $B$, this process can be described by a Langevin equation for the fraction $x$ of individuals carrying allele $B$,

$$\dot{x}(t) = f(t)x(t)[1 - x(t)] + \mu_0[1 - 2x(t)] + \chi_x(t), \quad (1)$$

where $\chi_x(t)$ are Gaussian random variables with zero mean and variance $\langle \chi_x(t)\chi_x(t')\rangle = [x(1-x)/N]\delta(t-t')$ and time is measured in units of generations. A typical stochastic trajectory is shown in Fig. 1(a). The process generates a statistical ensemble described by the allele frequency distribution $p(x, t)$; averages with respect to this ensemble are denoted by $\langle \ldots \rangle$.

Equation (1) is equivalent to Kimura's diffusion equation [4]

$$\dot{p}(x,t) = \frac{1}{2N}\nabla^2 x(1-x)p(x,t) - f(t)\nabla x(1-x)p(x,t) - \mu_0 \nabla(1-2x)p(x,t). \quad (2)$$

There are three evolutionary forces associated with three different time scales: (a) Selection acts via differences in fitness (i.e., subpopulation growth rate) $f(t) \equiv f_A(t) - f_B(t)$. For time-independent $f(t) = \sigma_0$, it leads to a deterministic change in allele frequencies over a characteristic time $1/\sigma_0$. (b) Mutations exchange alleles in an individual at a rate $\mu_0$;, i.e., the average time between mutations is of order $1/\mu_0$. (c) Reproductive fluctuations (referred to as genetic drift) are the stochasticity in an individual's number of offspring and change allele frequencies diffusively over a typical time of $\tau_0$ generations. This scale separates micro- and macro-evolution. The diffusion process ends with a deleterious or advantageous substitution, where the less fit or the fitter allele is established in the entire population (i.e., $x = 0$ or $x = 1$). For neutral evolution ($\sigma_0 = 0$), the diffusion time is twice the so-called effective population size, $\tau_0 = 2N$ [3]. Selection shortens $\tau_0$, with the asymptotics $\tau_0 \simeq 1/\sigma_0$ (up to log corrections) for strong selection ($\sigma_0 N \gg 1$).

Much of the statistical physics literature on evolution, in particular, on the celebrated quasispecies theory [5], implicitly or explicitly assumes populations of effectively infinite size, where reproductive fluctuations can be neglected. This regime is given by the ratio of time scales $\mu_0 N \gg 1$. However, most biological systems are in the opposite regime of low mutation rates in finite populations, $\mu_0 N \ll 1$, which is the subject of population genetics and is discussed in this Letter [6]. Macro-evolution of a genomic locus is then a sequence of substitutions interspersed with occasional coexistence periods (called polymorphisms) where both alleles reach substantial frequencies; see Fig. 1(d). The average polymorphism lifetime $\tau_0$ is much shorter than the average time between substitutions, leading to a dynamical separation that will be important below. The process eventually reaches an equilibrium state with constant fitness and detailed balance: every advantageous change just repairs a previous deleterious one; their respective numbers are equal.

 

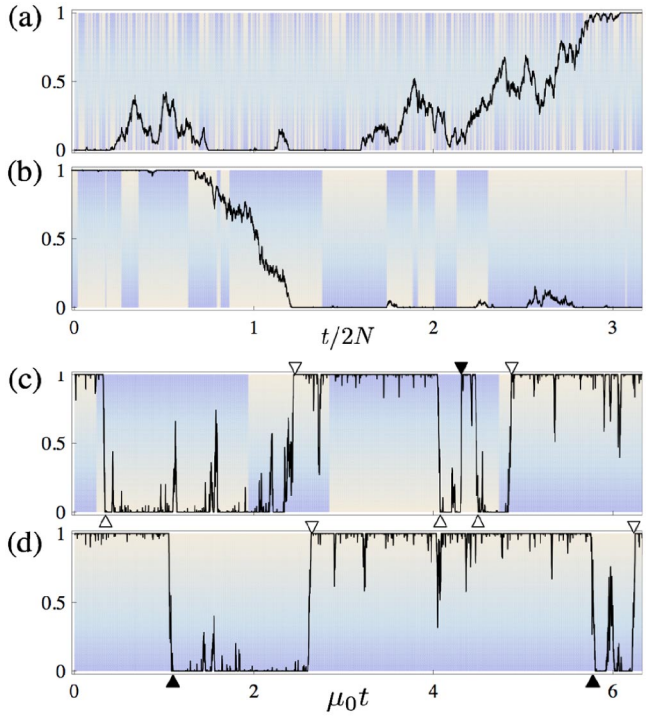


FIG. 1 (color online). Evolution under fluctuating selection, mutations, and genetic drift. One trajectory $x(t)$ of the stochastic process (1) with scaled parameters $2N\mu_0 = 0.02$ and $2N\sigma_0 = 2$, the direction of selection is indicated by shading (the fitter allele appears unshaded). (a) Selection with micro-evolutionary (annealed) fluctuations ($1/\gamma_0 \ll \tau_0$): Evolution can be described as a homogeneous diffusion process. Time is shown in units of $2N$. (b) Selection with intermediate fluctuations ($1/\gamma_0 \sim \tau_0$): The diffusion picture breaks down due to large frequency changes within one correlation interval of selection. (c) Selection with macro-evolutionary (quenched) fluctuations ($1/\gamma_0 \gg \tau_0$): The process reaches a nonequilibrium, time-irreversible stationary state with adaptations as surplus of advantageous substitutions (△) over deleterious ones (▲). Time is now shown in coarse-grained units $1/\mu_0$. (d) Time-independent selection ($\gamma_0 = 0$): The process reaches an evolutionary equilibrium with an equal number of advantageous and deleterious substitutions, hence without adaptations.

It is clear, however, that evolution under constant selection does not adequately represent the environment of most biological systems. Fitness effects vary in time and space, and these variations are the cause of phenotypic adaptations. To capture this effect quantitatively, selection itself should be regarded as a random force in the evolution Eqs. (1) and (2), as discussed by Wright, Kimura, Ohta, Gillespie, and others [7–13]. Here, we use the simplest model of fluctuating selection, where the time-dependent selection coefficient in (1) and (2) takes the form $f(t) = \sigma_0\eta(t)$ with a constant magnitude $\sigma_0$ and a random direction $\eta(t) = \pm 1$, which follows a Poisson process with rate $\gamma_0$. Selection fluctuations must be distinguished from reproductive fluctuations. They define a second statistical ensemble, averages with respect to which are denoted by overbars. It is given by the expectation value and covariance of the variables $\eta(t)$,

$$\overline{\eta(t)} = 0, \qquad \overline{\eta(t)\eta(t')} = e^{-2\gamma_0|t-t'|}, \tag{3}$$

and introduces a further scale, the correlation time $1/\gamma_0$.

The classical work on fluctuating selection was mainly aimed at describing micro-evolutionary effects such as seasonality or frequency-dependent selection, where correlation times do not exceed the diffusive time scale, $1/\gamma_0 \lesssim \tau_0$ [12]. In the language of statistical physics, the selection coefficients are annealed random variables, and averaging over this ensemble produces an effective diffusion equation for allele frequencies [8,10,12,14,15]. However, there is also recent evidence for selection fluctuations on the much larger mutation time scale, $1/\gamma_0 \sim 1/\mu_0 \gg \tau_0$, driving macro-evolutionary changes observed in a cross-species comparison of fly genomes [16]. These are probably caused mainly by epistasis, i.e., fitness interactions between genomic sites. Such interactions are probably frequent [17], in particular, in regulatory DNA [18,19]. Macro-evolutionary selection fluctuations have also been proposed as an explanation for anomalies of the molecular clock [13]. Such fluctuations are a quenched random process, for which diffusion theory breaks down. Instead, the joint statistics of selection and reproduction becomes similar to a system with thermal fluctuations and frozen disorder in physics.

In this Letter, we develop the theory of fluctuating selection with arbitrary correlation times. Genomic evolution under this type of selection is measured by time-dependent allele frequency correlations averaged over both the reproduction and the selection ensemble,

$$G(t_2 - t_1) = \overline{\langle x(t_2)[1 - x(t_1)]\rangle}, \tag{4}$$

which can be compared to data averages over a large number of genomic loci assumed to evolve independently. The variation within a population is given by $\bar{\Delta} \equiv G(\tau = 0)$, the regime $G(t) \simeq \bar{u}t$ (for $\tau_0 \ll t \ll 1/\mu$) describes the divergence between populations increasing with their evolutionary distance. Depending on the ratio between fitness correlation time $1/\gamma_0$ and diffusion time $\tau_0$, our theory identifies three evolutionary regimes as shown in Fig. 1(a)–1(c): (i) micro-evolutionary fluctuations ($1/\gamma_0 \ll \tau_0$), where classical diffusion theory holds, (ii) the crossover regime ($1/\gamma_0 \sim \tau_0$), where the substitution rate $\bar{u}$ is shown to have a maximum (as numerically observed in [13]), and macro-evolutionary fluctuations ($1/\gamma_0 \gg \tau_0$) leading to adaptive substitutions, defined as surplus of advantageous over deleterious substitutions.

To establish our results, we first recall the classical theory of Kimura's diffusion equation [4]. For time-independent selection, Eq. (2) can be written in the form of a continuity equation [20], $\dot{p} = -\nabla J p$ with the current operator $J = -\nabla x(1-x) + \sigma x(1-x) + \mu(1-2x)$. Here, we measure time in units of the neutral diffusion time $2N$ and introduce the accordingly scaled parameters $\mu \equiv 2N\mu_0$ and $\sigma = 2N\sigma_0$. From now on, we assume

$\mu \ll 1$. Two normalized and linearly independent distributions $p_\alpha$ ($\alpha = \pm 1$) with constant current $Jp_a = u_a$ can be defined by the boundary conditions that $p_+(x)$ remain finite at $x = 1$ and $p_-(x)$ remain finite at $x = 0$. These population states describe the restricted ensembles of paths $x(t)$ starting from $x = 0$ by an initial mutation $A \to B$ and from $x = 1$ by an initial mutation $B \to A$, respectively, and ending by fixation of either allele. The population variance in these states is given by the moments $\Delta_a \equiv \int x(1-x)p_a(x)dx$ (which can be expressed in terms of hypergeometric functions) [21]. Furthermore, the current values $u_a = a\mu\sigma/(1 - e^{-a\sigma})$ define the well-known Kimura-Ohta rates for substitutions $A \to B$ and $B \to A$ [22,23]. A generic normalized stationary solution $p(x)$ of (2) takes the form $p(x) = \lambda p_+(x) + (1-\lambda)p_-(x)$ with $0 \le \lambda \le 1$. The equilibrium distribution $p_{\rm eq}(x)$ is the unique stationary solution with $Jp_{\rm eq}(x) = 0$, i.e., with detailed balance between forward and backward substitutions, which determines $\lambda_{\rm eq} = 1/(1 + e^{\sigma}) = u_-/(u_+ + u_-)$.

Diffusion theory can be extended to micro-evolutionary selection fluctuations ($1/\gamma \ll \tau$ in terms of the rescaled parameters $\gamma \equiv 2N\gamma_0$ and $\tau = \tau_0/2N$). In this annealed approximation, fluctuating selection generates additional frequency-dependent diffusion and transport terms in the Fokker-Planck current, $J_s(g) = g\nabla x^2(1-x)^2 + gx(1-x)(1-2x)$, which depend on the effective strength $g \equiv \sigma^2/2\gamma$ [14,15]. For sufficiently large frequencies ($g \ll 1$), evolution is asymptotically neutral with stationary variance $\bar{\Delta} = \mu$ and substitution rate $\bar{u} = \mu$. Selection effects set on for $g \sim 1$, decreasing $\bar{\Delta}$ and increasing substitutions, with

$$\frac{\bar{u}}{\mu} \simeq \frac{g}{2\log g} \quad \text{for } g \gg 1. \tag{5}$$

We have recently developed a complementary analytical approach, which is valid for macro-evolutionary selection fluctuations ($1/\gamma \gg \tau$) [16]. From an arbitrary initial distribution $p(x, t = 0)$, the frequency distributions describing forward and backward paths reach their stationary shapes $p_+(x)$ and $p_-(x)$ within an initial time regime of order $\tau_0$ generations, provided the selection is quenched on these time scales. For larger values of time, the distribution takes the quasistationary form $p(x, t) = \lambda(t)p_+(x) + [1 - \lambda(t)]p_-(x)$ up to correction terms of order $\exp(-t/\tau)$ in rescaled time units, the time-dependent coefficient $\lambda(t)$ describing the long-term dynamics of substitutions. Defining $h(t) \equiv 1 - 2\lambda(t)$, $v \equiv u_+ + u_-$, and $w \equiv u_+ - u_-$, we obtain a Langevin equation with additive noise, $\dot{h}(t) = -vh + w\eta(t)$, which can be solved exactly. A key observable of the system is the stationary allele frequency bias

$$\bar{H} \equiv \lim_{t\to\infty} \overline{h(t)\eta(t)} = \frac{w}{2\gamma + v}, \tag{6}$$

which measures the degree of adaptation to the momentary direction of selection ($0 < H < 1$). This also determines the stationary variation within the population

$$\bar{\Delta}(\sigma, \gamma, \mu) = \frac{1}{2}(1 - \bar{H})\Delta_+ + \frac{1}{2}(1 + \bar{H})\Delta_- \tag{7}$$

and the stationary substitution rate

$$\bar{u}(\sigma, \gamma, \mu) = \frac{1}{2}(1 - \bar{H})u_+ + \frac{1}{2}(1 + \bar{H})u_-. \tag{8}$$

$\bar{H}$ is decreased, $\bar{\Delta}$ and $\bar{u}$ are increased with respect to equilibrium, which emerges in the limit $\gamma \to 0$.

The crossover from micro- to macro-evolutionary selection fluctuations as a function of $\gamma$ is shown in Fig. 2 by numerical simulations for $\bar{H}$, $\bar{\Delta}$, and $\bar{u}$ together with the results of diffusion theory and quenched selection theory. The annealed approximation is seen to break down for $\gamma \lesssim 1/\tau$, where memory effects of selection can no longer be neglected and frequency changes of order one occur within one correlation interval of selection cf. Figure 1(b). Conversely, the quenched approximation breaks down for $\gamma \gtrsim 1/\tau$, where fitness changes occur during one substitution process. There is currently no closed solution for generic values of $\gamma$, but we can interpolate between both asymptotic solutions to obtain the behavior in the crossover region $\gamma \sim 1/\tau$. The substitution rate $\bar{u}$, in particular, changes differently with $\gamma$ in the micro- and macro-evolutionary regime and, hence, must have a maximum $\bar{u}_{\max}$ at the crossover frequency $\gamma_{\max}$. Equating (5) and (8), we have

$$\gamma_{\max} \simeq \frac{\sigma}{2\log\sigma}\bar{u}_{\max} \simeq \frac{u_+}{2} \tag{9}$$

for strong selection and low mutation rate ($\sigma \gg 1$ and $\mu \ll 1$, which implies $u_+ \simeq \sigma\mu$ and $u_- \simeq e^{-\sigma}\mu$), where the selective enhancement of the substitution rate is most pronounced. The time scale $\tau = 1/\gamma_{\max} \simeq 2\log\sigma/\sigma$ is indeed proportional to the average lifetime of polymorphisms in this regime [24]. The enhancement of $\bar{u}$ can be observed in a population bottleneck: ecological selection fluctuations, which are micro-evolutionary and, hence, well balanced by a large population, can cause a dramatic increase of the substitution rate to values above the neutral rate during a population bottleneck while $N$ is temporarily reduced. (This increase goes beyond the known effect of a bottleneck with stationary selection, where $u$ reaches near-neutral values by temporary removal of selective constraint.) The degree of adaptation $\bar{H}$ is seen to be exponentially small in the micro-evolutionary regime (and strictly 0 in the diffusion approximation). Adaptation emerges at the crossover point $\gamma \sim 1/\tau$, increases with decreasing $\gamma$ in the macro-evolutionary regime, and reaches saturation ($\bar{H} \approx 1$) for fluctuations on the mutational time scale ($\gamma \gtrsim v \simeq \sigma\mu$ for $\sigma \gg 1$). Thus, it is the long-term fluctuations of selection that are most relevant for the adaptive diversification between species, and these can be detected by genomic cross-species analysis [16]. This regime clearly shows the time-irreversibility of the driven evolution process: adaptations always take place after the selection

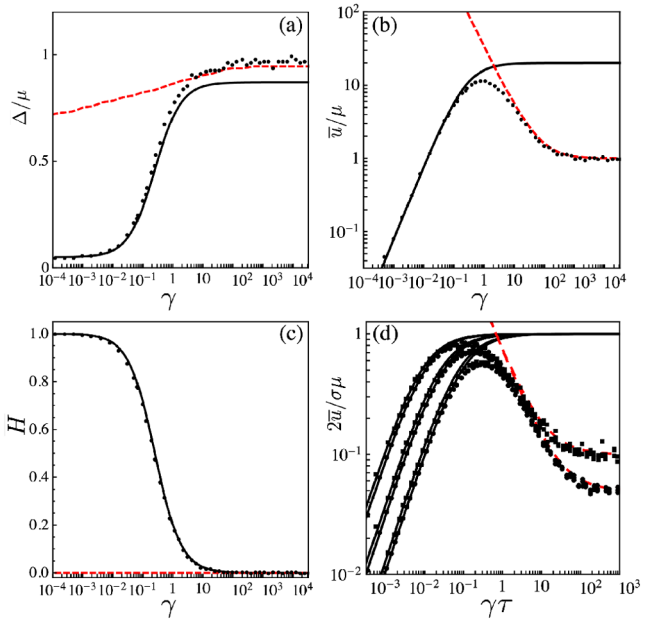


FIG. 2 (color online). (a) Population variance $\bar{\Delta}$, (b) substitution rate $\bar{u}$, and (c) degree of adaptation $\bar{H}$ at stationarity under fluctuating selection, shown as a function of the fluctuation rate $\gamma$ for $\sigma = 40$. Numerical results (data points) together with analytical results of diffusion theory [14] (dashed lines) and of quenched fluctuation theory [16] (full lines) valid in the micro-evolutionary regime ($\gamma \gg 1/\tau$) and the macro-evolutionary regime ($\gamma \ll 1/\tau$), respectively. (d) Crossover scaling: $2\bar{u}/\sigma\mu$ is plotted against $\gamma\tau$ for $\mu = 0.0028, 0.0084, 0.025$ (data points, from top to bottom) and $\sigma = 20, 40$ (squares, points), showing a data collapse in the crossover region $\gamma \sim 1/\tau$. Diffusion theory (dashed lines) and quenched theory (solid lines) are seen to be valid for $\gamma \gtrsim 1/\tau$ and $\gamma \lesssim 1/\tau$, respectively.

switch causing them, a clear difference to the time-reversible macro-evolution under constant selection cf. Figure 1(c). Evidence that genome evolution is off equilibrium (i.e., detailed balance is broken) has been found previously for neutral four-allele mutation processes with unequal rates $\mu_{AB}$ [25]. With the selection characteristics inferred for Drosophila [16], our analysis shows that detailed balance is strongly broken on a genome-wide scale, independently of the form of the neutral rate matrix.

In conclusion, we establish the statistics of genomic evolution driven by two stochastic forces: "diffusive" reproduction together with fitness fluctuations, which change from annealed to quenched randomness depending on their correlation time. For sufficient levels of selection ($\sigma = 2N\sigma_0 \gg 1$), there are three different characteristic selection frequencies governing the onset of selection effects ($\gamma \sim \sigma^2$), the crossover from annealed to quenched fluctuations ($\gamma \sim \sigma/2\log\sigma$), and the onset of adaptive evolution ($\gamma \sim \sigma\mu$), respectively. Our results lay the ground for a more powerful analysis of sequence data to infer adaptation based on a dynamic model of selection. Although the specific process we discuss here is solvable by elementary means, even close variants are more difficult and provide a rich playground for advanced methods of disordered systems physics. There is indeed evidence for genome-wide fitness fluctuations at individual sites, which probably signal both external forces and interactions with other sites [16]. Could evolutionary innovations, then, be avalanches of strongly correlated genomic changes, where each substitution is not only a response to its local fitness function but also a force shaping the fitness functions of other genomic positions? These ideas have been a long-standing theme of phenotypic evolution models [26] and may now become testable against genomic data. The corresponding sequence evolution models, however, have yet to be developed.

This work was supported in part by DFG Grant SFB 680 and by the National Science Foundation under Grant No. PHY05-51164 to the KITP (UC Santa Barbara), whose hospitality we gratefully acknowledge.

[1] M. Kreitman, Annu. Rev. Genomics Hum. Genet. **01**, 539 (2000).
[2] S. Wright, Genetics **16**, 97 (1931).
[3] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).
[4] M. Kimura, J. Appl. Prob. **1**, 177 (1964).
[5] M. Eigen, Naturwissenschaften **58**, 465 (1971); For a physicist's review, see, L. Peliti, arXiv:cond-mat/9712027.
[6] The crossover to the quasispecies regime in the presence of fluctuating selection will be discussed in G. J. Szollosi, V. Mustonen, and M. Lässig (to be published).
[7] S. Wright, Evolution (Lawrence, Kansas) **2**, 279 (1948).
[8] M. Kimura, Genetics **39**, 280 (1954).
[9] J. H. Gillespie, Theor. Popul. Biol. **3**, 241 (1972).
[10] T. Ohta, Genet. Res. **19**, 33 (1972).
[11] T. Ohta, J. Mol. Evol. **1**, 305 (1972).
[12] J. H. Gillespie, *The Causes of Molecular Evolution* (Oxford University Press, Oxford, 1991).
[13] J. H. Gillespie, Genetics **134**, 971 (1993).
[14] N. Takahata, K. Ishii, and H. Matsuda, Proc. Natl. Acad. Sci. U.S.A. **72**, 4541 (1975).
[15] N. Takahata and M. Kimura, Proc. Natl. Acad. Sci. U.S.A. **76**, 5813 (1979).
[16] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U.S.A. **104**, 2277 (2007).
[17] D. M. Weinreich, R. A. Watson, and L. Chao, Evolution (Lawrence, Kansas) **59**, 1165 (2005).
[18] J. Berg, S. Willmann, and M. Lässig, BMC Evolutionary Biology **4**, 42 (2004).
[19] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U.S.A. **102**, 15936 (2005).
[20] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2001).
[21] S. A. Sawyer and D. L. Hartl, Genetics **132**, 1161 (1992).
[22] M. Kimura, Genetics **47**, 713 (1962).
[23] M. Kimura and T. Ohta, Genetics **61**, 763 (1969).
[24] M. Desai and D. Fisher, Genetics **176**, 1759 (2007).
[25] P. Arndt and T. Hwa, Bioinformatics **20**, 1482 (2004).
[26] R. V. Solé *et al.*, Trends Ecol. Evol. **14**, 156 (1999).