Does fitness increase in adaptive evolution?

Ville Mustonen * **and Michael Lässig** *

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

Natural selection favors fitter variants in a population, but actual evolutionary processes may decrease fitness by mutations and genetic drift. How is the stochastic evolution of molecular biological systems shaped by natural selection? Here we derive a theorem on the fitness flux in a population, defined as the selective effect of its genotype frequency changes. The fitness flux theorem generalizes Fisher's fundamental theorem of natural selection to evolutionary processes including mutations, genetic drift, and time-dependent selection. It shows that a generic state of populations is adaptive evolution: there is a positive fitness flux resulting from a surplus of beneficial over deleterious changes. In particular, stationary nonequilibrium evolution processes are predicted to be adaptive. Under specific nonstationary conditions, notably during a decrease in population size, the average fitness flux can become negative. We show that these predictions are in accordance with experiments in bacteria and bacteriophages, and with genomic data in Drosophila. Our analysis establishes fitness flux as a universal measure of adaptation in molecular evolution.

Adaptive evolution │ fitness landscapes │ Fundamental theorem of natural selection │ fluctuation theorems in statistical physics

Adaptive processes have become centerstage in molecular evolutionary biology. Massive sequencing of populations opens unprecedented opportunities to trace the genomic basis of adaptation in population-genetic studies within and across species, as well as in time series of evolution experiments. Many different methods are used to infer natural selection from such data, however, their results lack a common gauge and are sometimes difficult to compare. This paper develops the notion of fitness flux in a population as a generic measure of adaptation applicable to molecular data and shows that fitness flux, not fitness, is the right variable to address the question of its title. A partial answer to this question turns out to be contained already in Fisher's fundamental theorem of natural selection [1]: in an evolution process governed by natural selection alone, any fitness difference within a population leads to adaptation. The fitness flux of this deterministic process equals the (additive) fitness variance in the population: the flux is positive when adaptation occurs and zero otherwise.

Generalizing this picture to realistic processes of molecular evolution has been a long-standing problem [3, 4, 5, 6, 7, 8]. The solution presented here involves a number of important conceptual steps. Molecular processes are always stochastic due to genetic drift and mutations, and we include these forces into a stochastic theory of fitness flux. Secondly, we extend the observation of this dynamics to the time scales of genomic data: populations are described by histories of genotype composition and demography which may extend beyond their coalescence time. Thirdly, natural selection itself is treated as dynamic on these time scales: we generalize static fitness landscapes, a concept introduced by Wright, to explicitly time-dependent models of selection referred to as fitness *seascapes*. The time dependence of selection reflects the changing ecology of a species. It has complex and opposing effects on adaptation [9, 10, 11, 12, 13, 14]: rapid fluctuations enhance the stochasticity of the evolutionary process and impede long-term adaptation [10], but persisting changes open windows of positive selection and are the very cause of adaptive evolution and fitness flux [12, 14]. The production of fitness flux by adaptive processes is a nonequilibrium phenomenon and does not necessarily imply any increase in fitness. Surprisingly, this production obeys a general theorem, which is the central result of this paper and gives new answers to the title question for stochastic evolution. The theorem and its proof establish an important conceptual link between evolutionary genetics and stochastic nonequilibrium thermodynamics. A close analogy of the fitness flux theorem to its deterministic counterpart, Fisher's theorem, emerges in the case of stationary evolution: the average fitness flux is positive in nonequilibrium stationary states – which we associate with adaptive evolution [14], and is zero at evolutionary equilibrium – where no adaptation occurs. The power of the fitness flux theorem will be demonstrated by a number of applications to evolution experiments in microbes and to cross-species genome comparison in flies.

Theory of Fitness Flux

Population histories and fitness flux. We start by introducing the notion of fitness flux and its relationship to fitness. Consider first the microevolution of a population containing a resident and a mutant genotype. The genotype evolution of the population can be described by a series of observations $\mathbf{x} = (x_0, x_1, \ldots, x_n)$ of mutant frequencies at successive times $(t_0, t_1, \ldots, t_n)$, which we refer to as the history of the population. Natural selection governs this process by a difference in reproductive rate of the mutant against the resident genotype, s, which may depend on their genotype frequencies and on time. Here we consider selection coefficients $s(x, t)$ with persistence times of many generations, which affect the evolution of genotype frequencies and can be measured, for example, by growth competition experiments in a microbial population. Measurements of population histories are assumed to be sufficiently dense so that the changes in selection as well as frequency changes $\Delta x_i = x_{i+1} - x_i$ in each time interval $\Delta t_i = t_{i+1} - t_i$ are small. The *fitness flux* can then be defined as a measure of adaptation at a given point of the population history: $\phi(t_i) = s(x_i, t_i)\Delta x_i/\Delta t_i$ is the product of the selection coefficient and the rate of frequency change. The cumulative fitness flux $\Phi(\mathbf{x}) = \sum_{i=0}^{n-1} \phi(t_i)\Delta t_i = \sum_{i=0}^{n-1} s(x_i, t_i)\Delta x_i$ measures the total adaptation of the entire population history [12]. This picture is easily extended to a space of k genotypes, where x and s become vectors with $k - 1$ independent components (for details, see *Methods*). Including all accessible genotypes in this space, not just those coexisting at a given point in time, makes it possible to describe processes over longer evolutionary time. Such processes may involve several genomic substitution events and fixed population states interspersed with polymorphic time intervals (in a fixed population state, one component of x is 1 and all others vanish). The genotype-based approach is useful for asexual populations or for compact genomic units such as transcription factor binding sites, where recombination can be neglected. Genotype-dependent selection models can also describe fitness interactions between genomic loci (epistasis) [15]. Our

---

Reserved for Publication Footnotes

results equally apply if x denotes a set of allele frequencies at individual genomic loci, a picture often used for recombining populations.

Measuring fitness flux. Population histories and fitness fluxes are increasingly accessible to experimental observation [16, 17, 18, 19]. As an example, consider a recent experiment describing the adaptation of a bacterial population to antibiotic stress [16]. This process involves five aminoacid substitutions in a specific protein, i.e., $2^5 = 32$ different genotypes and $5! = 120$ different population histories differing in the order of these substitutions. The selection coefficient of each substitution in each genetic background can be measured, and the cumulative fitness flux of a population history is simply the sum of the selection coefficients of its substitutions in their specified order. More generally, polymorphism and substitution data from a growing number of sequenced genomes can be used to estimate rate and selection coefficients of sequence changes. Such estimates provide copious data on fitness fluxes and adaptation on macro-evolutionary scales, albeit averaged over time and genomic sequence categories. It can be shown, for example, that the well-known McDonald-Kreitman method [22] is a test not merely for positive selection, but for the positivity of fitness flux [12].

Fitness land- and seascapes. A particularly intuitive picture of fitness flux emerges if the selection coefficients are the gradient of a time-independent *fitness landscape*, $s(x) = \nabla F(x)$, as in the microbial experiment just mentioned. If the reproductive rates of all genotypes are frequency-independent constants, the function $F(x)$ equals the mean population fitness; see *Supporting Information (SI) Text*. In any fitness landscape, the frequency change Δx_i in each time interval between successive measurements results in a fitness change $\Delta x_i \nabla F(x_i)$, as illustrated in Fig. 1(a). Hence, the cumulative fitness flux $\Phi(\mathbf{x}) = \sum_{i=0}^{i-1} \Delta x_i \nabla F(x_i)$ of the entire population history is simply the fitness change between the initial and the final population, $\Phi(\mathbf{x}) = F(x_n) - F(x_0)$. This picture can be extended to evolution in a *fitness seascape* $F(x, t)$ describing time-dependent selection coefficients $s(x, t) = \nabla F(x, t)$, as illustrated in Fig. 1(b). However, the resulting fitness flux $\Phi(\mathbf{x}) = \sum_{i=0}^{i-1} \Delta x_i \nabla F(x_i, t_i)$ is no longer related to any fitness difference between the initial and the final population. In the example of Fig. 1(b), the initial state is fitter than the final state in the original landscape, $F(x_0, t_0) > F(x_n, t_0)$, but the roles of both states are reversed in the final landscape, $F(x_n, t_n) > F(x_0, t_n)$. The example shows that fitness increase or decrease of a population history cannot even be defined in an unambiguous way under time-dependent selection and, hence, cannot serve as a universal measure of adaptation. The same is true in more general cases in which the selection coefficients $s(x, t)$ cannot be expressed as gradient of any fitness function $F(x, t)$. An example is cyclic selective advantage between three or more genotypes as in the well-known rock-paper-scissors game, which is shown in Fig. 1(c) and has been observed, for example, in bacterial and lizard populations [20, 21]. We include these cases into the notion of fitness seascapes, having in mind that a seascape can have water currents unrelated to the height pattern of its waves. Fitness flux ϕ and cumulative flux Φ do not suffer these ambiguities of definition and are well-defined measures of a population's response to selection pressures in any fitness seascape. Positive values of ϕ, or an increase of Φ, signal adaptation in a population history. But when does Φ increase?

Deterministic vs. stochastic evolution, Fisher's theorem. The first to address the increase of Φ and to establish a measure of adaptation different from fitness was R.A. Fisher in the context of the fundamental theorem [1], whose rationale has been elucidated decades later by Price and Ewens [23, 24]. Fisher's *partial* rate of fitness increase due to natural selection [23] is just the fitness flux ϕ defined for allele frequencies at genomic loci (see *SI Text*). The fundamental theorem equates this flux to the additive genetic fitness variance in the popu-

lation. As long as there are any fitness differences between coexisting alleles, this flux is positive, i.e., the cumulative flux Φ increases, and there is adaptation: Fisher's populations move uphill on fitness landscapes. However, the theorem is valid only under the restrictive assumption that evolution is deterministic and dominated by natural selection, which excludes mutations and genetic drift. Hence, Fisher's theorem applies only to microevolution of existing alleles in the limit of large populations under strong selection (see also the discussion below).

To describe more realistic processes of molecular evolution over longer evolutionary times, we must include mutations and genetic drift in our scenario. These stochastic evolutionary forces counteract natural selection and invalidate Fisher's theorem. The fixation of slightly deleterious mutations and mutation load are well-known examples for such effects. Thus, the fitness flux of a population can become negative: stochastically evolving populations can move downhill on fitness landscapes [25]. Of course, stochastic evolutionary theory no longer addresses genotype frequencies and fitness flux for a single population, but averages over ensembles of populations and their histories. The theorem to be established below is an identity for specific averages of the cumulative fitness flux.

Evolutionary equilibria and fitness. We first show how fitness and fitness flux can be evaluated from an ensemble of stochastically evolving populations with a time-dependent probability distribution $P(x, t)$ of genotype frequencies. The evolution of this distribution is determined by the conditional probabilities $G(x', t', x, t)$ of the transition from intitial genotype frequencies x at time t to frequencies x' at a later time t'. We define evolutionary equilibrium as a stationary (time-independent) distribution $P_{eq}(x)$ which satisfies the so-called *detailed balance* condition $G(x', t', x, t)P_{eq}(x) = G(x, t', x', t)P_{eq}(x')$ for arbitrary times and frequencies: at equilibrium, the probability of any evolutionary transition equals the probability of the reverse transition. In the absence of selection, the mutation-drift process has an equilibrium $P_0(x)$, if mutation rates are time-independent and satisfy mild additional conditions (see *Methods*). Given a neutral equilibrium, the full process in an arbitrary time-independent fitness landscape $F(x)$ has an equilibrium of remarkably simple form,

$$P_{eq}(x) = P_0(x)\, e^{NF(x)}, \qquad \qquad \textbf{[1]}$$

where N is the effective population size (see *Methods*). Kimura's "U-shaped" distributions for a single two-allele locus at neutrality and under directional selection [25] are a classic example of evolutionary equilibria and of the fitness relation (1). Various forms of this relation have been used to infer scaled fitness landscapes $NF(x)$ from histograms of the distributions $P_0(x)$ and $P_{eq}(x)$ obtained from genomic data [26, 27] (for a recent review, see [14]). The fitness relation has an obvious information-theoretic interpretation [28]: the function $NF(x)$ is the relative log-likelihood of the distributions $P_{eq}(x)$ and $P_0(x)$, and its expectation value $N\langle F \rangle = \sum_x NF(x)P_{eq}(x)$ equals the relative (Kullback-Leibler) entropy of these distributions, $N\langle F \rangle = H(P_{eq}|P_0)$.

Reverse histories and fitness flux. Equilibrium may be rather the exception than the rule in the evolution of biological systems. A generic process evolves from arbitrary initial conditions in an arbitrary fitness seascape $F(x, t)$, generating a time-dependent distribution $P(x, t)$ that may be far from equilibrium. This distribution can still be compared to the neutral equilibrium distribution $P_0(x)$ by the relative log-likelihood $\mathcal{H}(x) = \log[P(x, t)/P_0(x)]$. We now derive a non-equilibrium identity similar to (1) for the probability distribution $\mathcal{P}(\mathbf{x})$ of population histories in a given time interval (t_0, t_n). We define for each history $\mathbf{x}$ a *reverse* history $\mathbf{x}^T = (x_0^T, \ldots, x_n^T)$ as shown in Fig. 1: Starting from the point $x_0^T = x_n$ at time t_0, the history $\mathbf{x}^T$ evolves through the population states of $\mathbf{x}$ in reverse order and each transition has the opposite fitness effect as the corresponding transition in the original history. Using methods originally developed
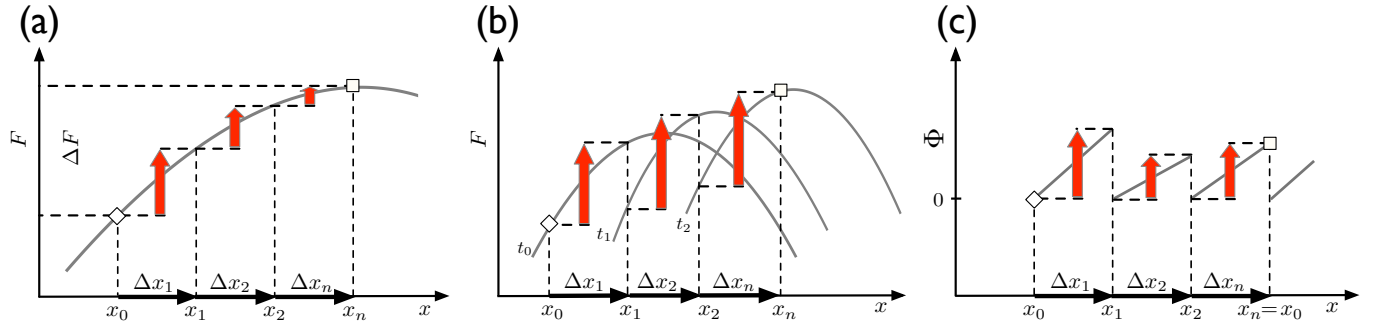
Fig. 1. Evolution in fitness landscapes and seascapes. The evolutionary history of a population is described by a series of genotype frequency states $\mathbf{x} = (x_0, \ldots, x_n)$ at times $(t_0, \ldots, t_n)$ (here $n = 3$). Evolutionary time increases between the initial state ($\diamond$) and the final state ($\square$). The cumulative fitness flux in each time interval (red arrows) is the product of the frequency change $\Delta x_i = x_{i+1} - x_i$ between successive states (black arrows) and the selection coefficient $s(x_i, t_i)$ of this change, the cumulative flux $\Phi(\mathbf{x})$ of the entire history is the sum of these terms. The reverse history $\mathbf{x}^T = (x_n, \ldots, x_0)$ evolves through the same states in reverse order, i.e., from the initial state ($\square$) to the final state ($\diamond$). Each transition has the opposite fitness effect as the corresponding transition of the original history, resulting in a cumulative fitness flux $\Phi(\mathbf{x}^T) = -\Phi(\mathbf{x})$ (the direction of all arrows is reversed).
(a) *Fitness landscape* $F(x)$, whose gradient defines time-independent selection coefficients $s(x) = \nabla F(x)$. A linear landscape corresponds to frequency-independent selection, a nonlinear landscape to frequency-dependent selection (the example shown here). The cumulative fitness flux $\Phi(\mathbf{x})$ of a population history measures the fitness difference $\Delta F = F(x_n) - F(x_0)$ between initial and final population. The function $F(x)$ is not always equal to the mean population fitness (as discussed in *SI Text*).
(b) *Fitness seascape* $F(x, t)$, whose gradient defines time-dependent selection coefficients $s(x, t) = \nabla F(x, t)$. The fitness flux of a population history is defined as before, but it no longer equals the fitness change between initial and final population.
(c) *Fitness seascape* with selection coefficients $s(x, t)$ not of gradient form. The example shows cyclic selective advantage as in the rock-paper-scissors game, i.e., each of the transitions from x_0 to x_1, from x_1 to x_2, etc., and from x_2 to $x_n = x_0$ involves a positive selection coefficient. The fitness flux is defined as before.

in nonequilibrium thermodynamics [29, 30, 31, 32], we can show that the probability of the reverse history is given by

$$\mathcal{P}(\mathbf{x}^T) = \mathcal{P}(\mathbf{x}) \, e^{-N\Phi(\mathbf{x}) + \Delta\mathcal{H}(\mathbf{x})}, \qquad [2]$$

where $\Phi(\mathbf{x})$ is the cumulative fitness flux and $\Delta\mathcal{H}(\mathbf{x}) = \mathcal{H}(x_n, t_n) - \mathcal{H}(x_0, t_0)$ is the difference in relative log likelihood between initial and final point of the original history (see *Methods*).

Fitness flux theorem. We can now state the theorem: An evolutionary process with mutations, genetic drift, and selection given by an arbitrary fitness seascape obeys the relation

$$\left\langle e^{-N\Phi + \Delta\mathcal{H}} \right\rangle = 1. \qquad [3]$$

It belongs to a set of relations known as fluctuation theorems, which have played an important role in the nonequilibrium statistical physics of mesoscopic systems over the past decade [29, 30, 31, 32]. The angular brackets denote an average over the probability distribution of population histories in a given time interval, $\langle f \rangle = \sum_{\mathbf{x}} f(\mathbf{x}) \mathcal{P}(\mathbf{x})$. Using eq. (2), we recognize this history average as the sum $\sum_{\mathbf{x}^T} \mathcal{P}(\mathbf{x}^T)$, which equals unity by normalization of the probability distribution of reverse histories. As an immediate consequence of eq. (3), the fitness flux has a lower bound for any such process,

$$N\langle \Phi \rangle \geq \Delta H, \qquad [4]$$

where $\Delta H = \langle \Delta\mathcal{H} \rangle = H(P(t)|P_0) - H(P(t_0)|P_0)$ is the relative entropy difference between initial and final frequency distribution.

Applications of the theorem

The relations (3) and (4) provide a unifying view on different scenarios of Darwinian evolution at the molecular level. We now discuss such scenarios and their relevance to experiments and genomic studies.

Evolutionary equilibrium. The equilibrium distribution (1) is the unique population ensemble in which the identity $N\Phi(\mathbf{x}) = N\Delta F(\mathbf{x}) = \Delta\mathcal{H}(\mathbf{x})$ holds for each population history. Numerical simulations of equilibrium populations shown in Fig. 2(a) illustrate the identical statistics of $N\Phi$ and $\Delta\mathcal{H}$ for individual histories

and for the resulting ensemble distributions. By eq. (2), the identity $N\Phi = \Delta\mathcal{H}$ implies that the probability of any history equals that of its reverse history, $\mathcal{P}(\mathbf{x}) = \mathcal{P}(\mathbf{x}^T)$, which expresses detailed balance of the equilibrium ensemble: beneficial substitutions of selection coefficient $s > 0$ occur at the same rate as deleterious substitutions of selection coefficient $-s$. Hence, the average fitness flux $\langle \phi \rangle$ vanishes, and the average fitness $\langle F \rangle$ remains constant. Evolutionary equilibrium has been realized, for example, in an experiment with bacteriophages [33]. Under stationary experimental conditions, the phage populations have been found to relax to constant fitness values which are independent of the previous history of the population and increase with population size, consistent with the behavior of the ensemble average $\langle F \rangle$ predicted by eq. (1). This "self-averaging" of individual populations suggests that the phage genome has a sufficient number of independently evolving loci. For the approach to equilibrium, the inequality (4) implies that the "free fitness" $NF - H$ always increases monotonically with time and is maximal at equilibrium [4, 34], which is an evolutionary analogue of Boltzmann's H theorem.

Nonequilibrium stationary states. In a generic population ensemble, the maximum principles of equilibrium [4, 34] and the detailed balance condition $N\Phi = \Delta\mathcal{H}$ for individual histories are violated. For any stationary nonequilibrium ensemble, the inequality (4) with $\Delta H = 0$ predicts that the average fitness flux $\langle \phi \rangle = d\langle \Phi \rangle / dt$ is a positive constant: populations steadily accumulate a surplus of beneficial over deleterious changes. Thus, adaptation takes place although the average fitness $\langle F \rangle$ remains constant. For most individual populations, the cumulative flux over a sufficiently long time interval Δt is close to the ensemble average, $\Phi \approx \langle \Phi \rangle = \phi \Delta t > 0$. The full theorem (3) shows that selection is overcome by genetic drift and $N\Phi < \Delta\mathcal{H}$ in an exponentially small subset of populations. Muller's ratchet, the scenario of degradation by an excess of deleterious substitutions, is thus very unlikely as a stationary process but requires, for example, a dwindling population size.

Stationary non-equilibrium evolution can be caused by time-dependent selection [14]. Here we consider a minimal stochastic fitness seascape defining independently fluctuating selection coeffi-
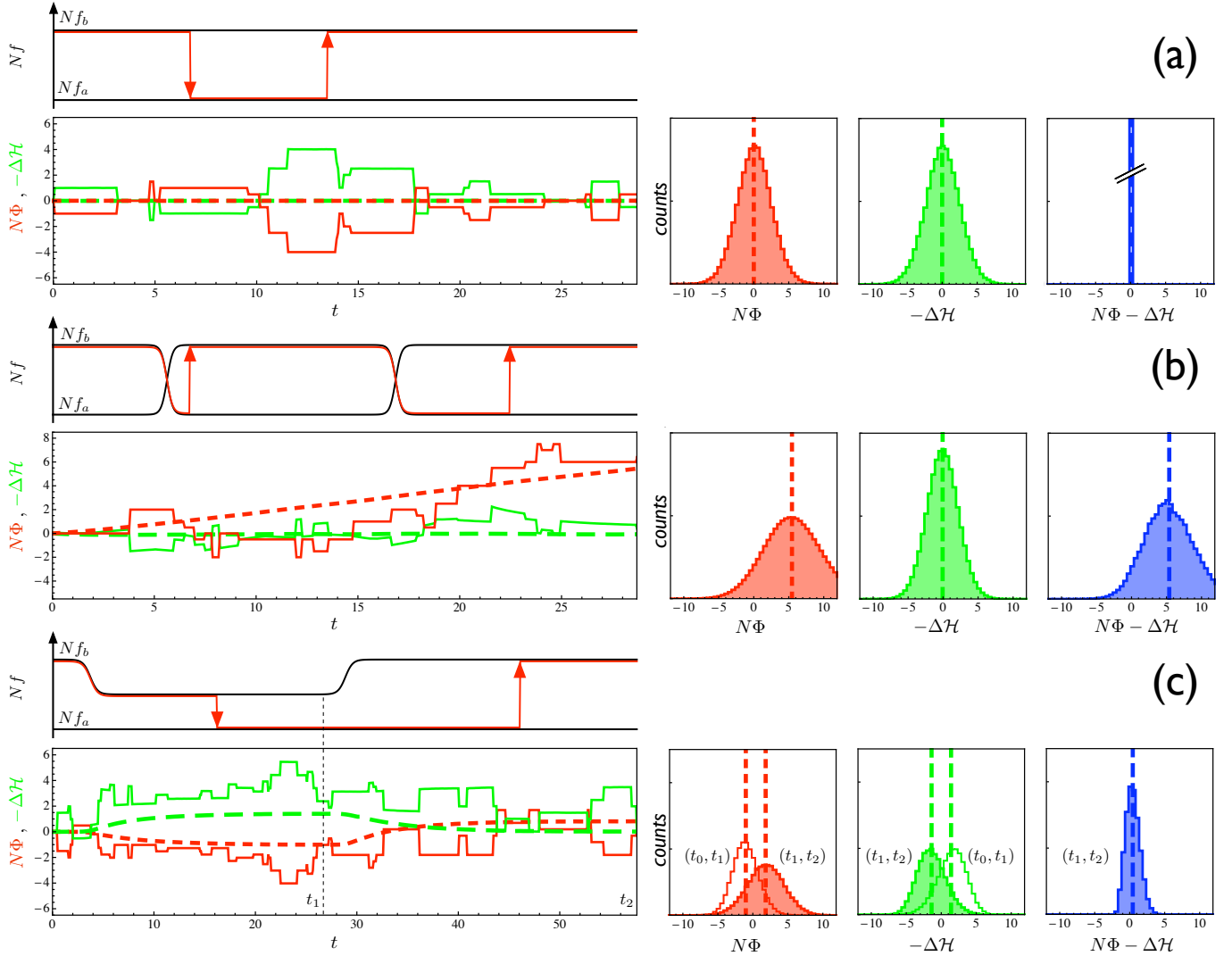
Fig. 2. Fitness evolution of genomic population histories under three different scenarios of selection and demography.
(a) *Evolutionary equilibrium* in a time-independent fitness landscape. Top diagram: Fitness evolution of a single two-allele genomic locus (schematic). The two alleles a, b have time-independent fitness values f_a, f_b (black lines). The mean population fitness of this locus (red line) evolves by a series of beneficial or deleterious substitutions (red arrows), which have selection coefficient $s = f_b - f_a$ and $-s$, respectively. This process obeys detailed balance, i.e., beneficial substitutions occur at the same rate as deleterious ones. Bottom left diagram: Fitness evolution of sequences with $L = 12$ independent two-allele loci, additive fitness and a uniform mutation rate μ per locus. Fitness flux $N\Phi$ (red lines) and the negative of the log likelihood change, $-\Delta\mathcal{H}$ (green lines), between initial and final population state in the interval $(0, t)$; time series of an individual history (solid lines) and average for an ensemble of 10^5 independently evolving populations (dashed lines). Each population history obeys the detailed balance relation $N\Phi = N\Delta F = \Delta\mathcal{H}$ (in the equilibrium ensemble, log likelihood $\mathcal{H}$ equals scaled fitness NF). Evolutionary time is measured in units of the neutral genomic mutation rate μL. For simulation details, see *SI Text*. Bottom right diagrams: Histograms of $N\Phi$ (red), $-\Delta\mathcal{H}$ (green), and $N\Phi - \Delta\mathcal{H}$ (blue) at a given time $t = 28.8$ for an ensemble of 10^5 populations, with averages marked by dashed vertical lines.
(b) *Nonequilibrium stationary state* in a stochastic fitness seascape. Diagrams as in (a). Selection coefficients $s(t) = f_b(t) - f_a(t)$ at individual genomic loci fluctuate between two values following a Poisson process, which generates independent selection histories at each locus(for details, see *SI Text*). The average cumulative fitness flux $N\langle\Phi\rangle$ increases with time at a constant positive rate. Most individual population histories have a flux $N\Phi$ close to this average, but there are rare drift-dominated histories with $N\Phi < -\Delta\mathcal{H}$.
(c) *Transitions between equilibria* under demographic changes. Diagrams as in (a). Population size first decreases from an initial value N_0 to a bottleneck value $N_b = N_0/2$, remains constant during the bottleneck, and later increases to the original value N_0. This process results in time-dependent scaled allele fitness values $N(t)f_a, N(t)f_b$ and selection coefficients $N(t)(f_b - f_a)$. The population decline generates a loss in scaled fitness, $\Delta_1 H = \Delta_1\langle NF \rangle < 0$, and a negative scaled fitness flux $N_0\langle\Phi_1\rangle < 0$ in the time interval $(0, t_1 = 26.6)$. The recovery restores the initial fitness, $\Delta_2 H = \Delta_2\langle NF \rangle = -\Delta_1 H > 0$, and generates a positive scaled fitness flux $N_0\langle\Phi_2\rangle > -N_0\langle\Phi_1\rangle$ in the time interval $(t_1, t_2 = 57.6)$ (for details, see *Methods* and *SI Text*).

cients at individual genomic loci. This process generates an ensemble of populations with joint histories of selection and genotype, which is shown in the numerical simulations of Fig. 2(b). In accordance with the fitness flux theorem, the cumulative flux $\langle\Phi\rangle$ in the stationary state increases at a constant positive rate $\langle\phi\rangle$: typical populations adapt to changing selection pressures by a surplus of beneficial over delete-

rious substitutions. Recent population-genetic studies of *Drosophila* genomes have shown evidence for adaptive evolution at a genome-wide level [35, 12, 36]. Positive fitness flux (with values $N\Phi$ of order 10 per genomic substitution [12]) has been inferred from joint estimates of rate and average selection coefficient of point mutations,

supporting the conclusion that a substantial fraction of the observed substitutions is adaptive at substantial levels of selection.

Demographic nonequilibrium. The fitness flux theorem also captures nonequilibrium processes generated by the demographical history of a population (see *Methods*). A simple example is a population bottleneck with a decline transition from equilibrium at an initial population size to a (near-)equilibrium at a lower population size, which is followed by a recovery transition to equilibrium at the initial population size. This is exactly the protocol of the bacteriophage experiment discussed above [33]. By eq. (1), the population decline leads to a loss in scaled fitness, $\Delta_1 H = \Delta_1 \langle NF \rangle < 0$, which is exactly compensated by the gain during recovery, $\Delta_2 H = \Delta_2 \langle NF \rangle = -\Delta_1 H$. The cumulative fitness flux $\langle \Phi_1 \rangle$ of the decline transition is allowed to become negative but is more than offset by the flux $\langle \Phi_2 \rangle$ of the recovery transition, such that the total flux $\langle \Phi_1 + \Phi_2 \rangle$ becomes positive. These predictions are confirmed by numerical simulations as shown in Fig. 2(c). More generally, any population whose population size changes periodically or fluctuates stochastically around some average will, over sufficiently long periods of time, acquire a positive cumulative flux.

Strong-selection limit and Fisher's theorem. Laboratory evolution experiments often involve very high selection pressures. In this regime, the probability of fitness-lowering frequency transitions becomes very small according to eq. (2), and only evolutionary histories with monotonically increasing fitness flux remain accessible to the system. This reduction of histories has been observed in the bacterial evolution experiment of ref. [16]. The evolutionary process over longer time intervals remains stochastic, since a new mutation appears by chance in an individual and its fate at small population frequencies remains governed by genetic drift. The remainder of its substitution process, however, becomes deterministic and follows a single most probable evolutionary history. In this limit, the fitness flux theorem contains Fisher's theorem (see *Methods*).

Conclusions

Here we have established fitness flux as a measure of adaptation in molecular evolution under time-dependent selection and population size, mutations, and genetic drift. The fitness flux theorem lays the statistical foundation of this measure and makes testable predictions in diverse contexts of experiment and genomics. These applications show that increase of Φ is an almost universal condition of biological systems: Positive contributions to the fitness flux arising from adaptive genotype changes accumulate over evolutionary periods of time, negative contributions are limited to time intervals with a systematic loss of adaptation ($\Delta H < 0$), which cannot occur continuously in viable populations. In this sense, fitness flux is a more fundamental characteristic of evolution than fitness, for which no comparable growth law holds.

Since Fisher speculated about a connection between the fundamental theorem and the second law of thermodynamics [1], several conceptual links between biological evolution and statistical thermodynamics have been discussed [37, 38]. Here we have demonstrated a generic connection away from equilibrium, which links molecular evolution and thermodynamics as stochastic processes driven by time-dependent forces. The fitness flux theorem shows that this dynamics follows common statistical principles in both fields.

Methods

Evolution equation and equilibrium distributions. In a sequence space of k different genotypes, the evolution of a population is described by time-dependent genotype frequencies $x(t) = (x^1, \ldots, x^k)(t)$ with the normalization $\sum_{\alpha=1}^{k} x^\alpha(t) = 1$, the evolution of an ensemble of populations by a time-dependent frequency distribution $P(x, t)$ with the normalization $\sum_x P(x, t) = 1$. For populations of large effective size N, frequencies can be approximated by continuum variables and the evolution takes the form of a diffusion equation [39],

$$\frac{\partial}{\partial t} P(x, t) = \sum_{\alpha, \beta=1}^{k-1} \frac{\partial}{\partial x^\alpha} \left[\frac{1}{N} \frac{\partial}{\partial x^\beta} g^{\alpha\beta}(x) - v^\alpha(x, t) \right] P(x, t).$$

[5]

The matrix $g^{\alpha\beta}(x)$ characterizes genetic drift, and the genotype evolution rates $v^\alpha(x, t) = m^\alpha(x) + \sum_\beta g^{\alpha\beta}(x) s_\beta(x, t)$ contain the contributions of mutations and of time-dependent selection with coefficients $s_\beta(x, t)$ (for details, see *SI Text*). A stationary neutral process with mutation rates μ_α^β satisfying the detailed balance conditions $\mu_\alpha^\beta / \mu_\beta^\alpha = p_0^\beta / p_0^\alpha$ can be shown to have an equilibrium distribution $P_0(x)$ (up to terms of order $\mu^2 N$, which are negligible for $\mu N \ll 1$, see *SI Text*). Inspection of the diffusion equation (5) then shows that the corresponding process in any time-independent fitness landscape $F(x)$ has an equilibrium distribution $P_{\text{eq}}(x)$ of the form (1).

Derivation of the fitness flux theorem (for details, see *SI Text*). Consider the ensemble of population histories $\mathbf{x} = (x_0, \ldots, x_n)$ with a sufficiently dense set of observation times $t_i = t_0 + i\delta$ ($i = 0, \ldots, n$). For a generic diffusion process of the form (5), this ensemble is characterized by a probability distribution $\mathcal{P}(\mathbf{x})$ of histories, which determines the frequency distribution $P(x_n, t_n) = \sum_{x_0, \ldots, x_{n-1}} \mathcal{P}(\mathbf{x})$ and has the normalization $\sum_{\mathbf{x}} \mathcal{P}(\mathbf{x}) = 1$. Using standard methods of statistical mechanics, the conditional probability of a history for given initial state, $G(\mathbf{x}) = \mathcal{P}(\mathbf{x})/P(x_0, t_0)$ can be shown to take the form $G(\mathbf{x}) = \exp\left[\delta \sum_{i=0}^{n-1} (\Delta x_i / \delta - v(\tilde{x}_i, \tilde{t}_i))^2 + c(\tilde{x}_i, \tilde{t}_i) \right]$, where we use the abbreviations $\Delta x_i = x_{i+1} - x_i$, $\tilde{x}_i = (x_i + x_{i+1})/2$, $\tilde{t}_i = t_i + \delta/2$, $\pi(x)^2 = \sum_{\alpha, \beta} \pi^\alpha(x) g_{\alpha\beta}(x) \pi^\beta(x)$ with coefficients $g_{\alpha\beta}$ given by the matrix inverse of $(g^{\alpha\beta})$, and the coefficient function $c(x, t)$ is determined by normalization. Now consider the reverse history $\mathbf{x}^T = (x_0^T, \ldots, x_n^T)$, which is given by $x_i^T = x_{n-i}$ ($i = 0, \ldots, n$) and, by definition, evolves in the time-reversed fitness landscape $F^T(x, t) = F(x, t_n - t + t_0)$ as shown in Fig. 1. Following refs. [30, 31, 32], we compare the conditional probabilities of original and reverse history,

$$\frac{G^T(\mathbf{x}^T)}{G(\mathbf{x})} = \frac{G_0^T(\mathbf{x}^T)}{G_0(\mathbf{x})} e^{-N\Phi(\mathbf{x})} = \frac{P_0(x_n)}{P_0(x_0)} e^{-N\Phi(\mathbf{x})}. \quad \textbf{[6]}$$

Inserting the definition of the relative log likelihood, $P(x, t) = P_0(x) e^{\mathcal{H}(x, t)}$, then yields Eq. (2), and the theorem (3) follows by summation over all histories.

Generalization to time-dependent population size. If we parametrize the population size in terms of a reference size, $N(t) = \zeta(t) N_0$, and define the scaled fitness flux as $N_0 \Phi(\mathbf{x}) = \sum_{i=0}^{n-1} N(t_i) \Delta x_i s(x_i, t_i)$, the identity (3) remains valid in very good approximation for low mutation rates. This is shown by an inhomogeneous rescaling of time by a factor $1/\zeta(t)$ in the evolution equation (5) (for details, see *SI Text*).

Fisher's theorem. In the strong-selection limit of the evolutionary process (5), the evolution of a polymorphic population is dominated by its most probable history (except close to fixation of one of the genotypes, $x < 1/Ns$). This history follows the deterministic evolution equation $dx_*^\alpha / dt = \sum_\beta g^{\alpha\beta}(x_*(t)) s_\beta(x_*(t), t)$. Hence, its fitness flux Φ_* increases at a rate $\phi_*(t) = d\Phi_*/dt =$

$(dx_*/dt)\, s(x_*(t),t) = [s(x_*(t),t)]^2$, which is the fitness variance in the population. Fisher's theorem is the projection of this identity from genotype frequencies to gene allele frequencies.

1. R. A. Fisher (ed. H. Bennet) The genetical theory of natural selection: a complete variorum edition. Oxford University Press, (2000).
2. G. R. Price, Selection and Covariance. *Nature* 227, 520-521 (1970).
3. M. Kimura, On the change of population fitness by natural selection. *Heretidy* 12(2), 145-167 (1958).
4. Y. Iwasa, Free fitness that always increases in evolution. *J theor Biol* 135, 265-281 (1988).
5. T. Nagylaki, Rate of evolution of a character without epistasis. *Proc Natl Acad Sci USA* 86, 1910-1913 (1989).
6. K. Sato, Y. Ito, T. Yomo, K. Kaneko, On the relation between fluctuation and response in biological systems. *Proc Natl Acad Sci USA* 100, 14086-14090 (2003).
7. M. O. Vlad, S.E. Szedlacsek, N. Pourmand, L.L. Cavalli-Sforza, P. Oefner, J. Ross, Fisher's theorems for multivariable, time- and space-dependent systems, with applications in population genetics and chemical kinetics. *Proc Natl Acad Sci USA* 102, 9848-9853 (2005).
8. S. H. Rice, A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evol Biol* 8, 262 (2008).
9. T. Ohta, Population size and rate of evolution. *J. Mol. Evol.* 1, 305-314 (1972).
10. J. H. Gillespie,Substitution Processes in Molecular Evolution. I. Uniform and Clustered Substitutions in a Haploid Model. *Genetics* 134, 971-981 (1993).
11. D. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309, 2075-2078 (2005).
12. V. Mustonen, M. Lässig, Adaptations to fluctuating selection in Drosophila. *Proc Natl Acad Sci USA* 104, 2277-2282 (2007).
13. V. Mustonen, M. Lässig, Molecular evolution under fitness fluctuations. *Phys Rev Lett* 100(10), 108101 (2008).
14. V. Mustonen, M. Lässig, From fitness landscapes to seascapes: Non-equilibrium dynamics of selection and adaptation. *Trends in Genetics* 25, 111-119 (2009).
15. D. M.Weinreich, R. A. Watson, L. Chao Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59, 1165Ð1174 (2005).
16. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111Ð114 (2006).
17. M. Desai, D. Fisher, A. Murray, The Speed of Evolution and Maintenance of Variation in Asexual Populations. *Current Biology*, 17:5, 385-394 (2007).
18. F. Poelwijk, D. Kiviet, D. Weinreich, S. Tans, Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445, 383-386 (2007).
19. J.A.G.M. de Visser, S.C. Park, J. Krug, Exploring the effect of sex on empirical fitness landscapes. *American Naturalist* 174: S15-S30 (2009).
20. B. Kerr, M. A. Riley, M. W. Feldman, B. J. M. Bohannan, Local dispersal promotes biodiversity in a real-life game of rockÐpaperÐscissors. *Nature* 418, 171-174, (2002).
21. B. Sinervo, C. M. Lively, The rock paper scissors game and the evolution of alternative male strategies. *Nature* 380, 240-243, (1996).
22. J. H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652 (1991).
23. G. R. Price, Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet. Lond.* 36, 129-140 (1972).
24. W. J. Ewens, An interpretation and proof of the fundamental theorem of natural selection. *Theor. Pop. Biol* 36, 167-180 (1989).
25. M .Kimura Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology* 20, 33-53 (1955).
26. A. L. Halpern, W. J. Bruno, Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol Biol Evol* 15, 910-917 (1998).
27. V. Mustonen, J. Kinney, C.G. Callan, M. Lässig, Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. *Proc Natl Acad Sci USA* 104, 2277-2282 (2008).
28. J. Berg, S. Willmann, M. Lässig, Adaptive evolution of transcription factor binding sites. *BMC Evol Biol* 4, 42 (2004).
29. C. Jarzynski, Nonequilibrium Equality for Free Energy Differences. *Phys Rev Lett* 78(14), 2690 (1997).
30. G. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E* 60(3), 2721 (1999).
31. U. Seifert, Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem. *Phys Rev Lett* 95(4), 040602 (2005).
32. V. Y. Chernyak, M. Chertkov, C. Jarzynski, Path-integral analysis of fluctuation theorems for general Langevin processes. *J. Stat. Mech.* P080001, (2006).
33. C. L. Burch, L. Chao, Evolution by small steps and rugged landscapes in the in the RNA Virus ϕ6. *Genetics*, 167, 559-567 (1999).
34. G. Sella, A. Hirsh, The application of statistical physics to evolutionary biology. *Proc Natl Acad Sci USA* 102: 9541-9546 (2005).
35. P. Andolfatto, Adaptive evolution of non-coding DNA in *Drosophila. Nature* 437, 1149-1152 (2005).
36. J. M. Macpherson, G. Sella, J. C. Davis, D. A. Petrov, Genomewide Spatial Correspondence Between Nonsynonymous Divergence and Neutral Polymorphism Reveals Extensive Adaptation in *Drosophila. Genetics* 177, 2083-2099 (2007).
37. L. Peliti, Introduction to the statistical theory of Darwinian evolution. cond-mat/9712027, (1997).
38. B. Drossel, Biological evolution and statistical physics. *Advances in Physics* 50(2), 209-295 (2001).
39. M. Kimura M *The neutral theory of molecular evolution.* Cambridge Univ. Press. (1983).