Opinion

Cell PRESS

From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation

Ville Mustonen and Michael Lässig

Institut für Theoretische Physik, Universität zu Köln, Zülpicherstraße 77, 50937 Köln, Germany

Evolution is a quest for innovation. Organisms adapt to changing natural selection by evolving new phenotypes. Can we read this dynamics in their genomes? Not every mutation under positive selection responds to a change in selection: beneficial changes also occur at evolutionary equilibrium, repairing previous deleterious changes and restoring existing functions. Adaptation, by contrast, is viewed here as a non-equilibrium phenomenon: the genomic response to time-dependent selection. Our approach extends the static concept of fitness landscapes to dynamic fitness seascapes. It shows that adaptation requires a surplus of beneficial substitutions over deleterious ones. Here, we focus on the evolution of yeast and *Drosophila* genomes, providing examples where adaptive evolution can and cannot be inferred, despite the presence of positive selection.

Evidence for selection in molecular evolution

Organismic evolution provides ubiquitous examples of adaptive evolution. Changes in the ecology of a population, such as migration to a new habitat, the conquest of a co-evolutionary niche or sympatric speciation, can produce phenotypic innovations driven by positive selection for change. At the molecular level, it has been an ongoing challenge to identify adaptive mutations, which take place in a sea of stochastic changes caused by genetic drift (i.e. by fluctuations in number of offspring in a finite population) [1–4].

In protein-coding DNA, most amino-acid-changing (non-synonymous) mutations are under negative selection [5], as shown by their substantially reduced substitution rates compared with synonymous mutations. A small subset of genes, including immune response and sex-related genes, show enhanced rates of non-synonymous substitutions [6], which indicate predominantly positive selection for change. Statistically more sensitive population-genetic tests based on substitutions and polymorphisms provide evidence for amino acid changes under positive selection in most genes [7–13]. At the same time, the functionality of non-coding DNA and the forces shaping its evolution are less clear. Regulatory elements encode biological information in a more fuzzy way than proteins. This can lead to considerable sequence divergence while the regulatory function is maintained, which makes adaptive evolution hard to detect. However, there is evidence for genome-wide positive selection of moderate strength in non-coding DNA [14–16], and complementary methods have identified selective sweeps under strong positive selection [17–20]. A sweep is the rapid fixation of a selected mutation, which also reduces the polymorphism of linked polymorphic loci in its neighborhood and, hence, becomes detectable by a contiguous interval of reduced diversity in the genome.

Do these genomic observations provide evidence for adaptive evolution as a response to changes in selection? A clear case can be made for selective sweeps: the observation of a sweep points to a strong change in selection at a genomic locus that has triggered its adaptive response soon after. However, selective sweeps are only a part of adaptive evolution. The majority of beneficial mutations identified by population-genetic tests have moderate selection coefficients. Inferring the cause of these changes is less straightforward because they take place in the context of genetic drift. A weakly beneficial change does not have to reflect a change in selection: it might just compensate a previously fixed weakly deleterious change without any change in selection. This has an important consequence, leading to the central thesis of this article: we should not equate positive selection and adaptation as is done implicitly in much of the literature. We need a notion of adaptation distinct from compensation, which is grounded on more precise models of phenotype and selection at the molecular level.

In the next section, we introduce a definition of adaptation as non-equilibrium response to changes in selection. This forces us to think of selection itself as a dynamical process, that is, to promote the static picture of fitness landscapes to dynamic fitness seascapes. The dynamical approach leads to a quantitative measure of adaptation called fitness flux, which counts the excess of beneficial over deleterious genomic change. The joint statistics of selection and genomic response can distinguish adaptation from compensation because it keeps track of temporal correlations in the direction of selection of subsequent mutations. As will become clear, this requires a departure from the well-known infinite-sites model [21,22], which neglects correlations between subsequent mutations, to a finite number of genomic sites governing a molecular phenotype [16,23]. In the following two sections, we illustrate the molecular concept of adaptation by two examples, the evolution of the yeast and fly genomes. The evolution of regulatory elements in *Saccharomyces* seems to be

Corresponding authors: Mustonen, V. (vimuston@thp.uni-koeln.de); Lässig, M. (lassig@thp.uni-koeln.de).

0168-9525/$ – see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2009.01.002 Available online xxxxxx

consistent with evolutionary equilibrium, a state in which we cannot infer adaptive evolution from genomic data. By contrast, the *Drosophila* genome shows evidence of a non-equilibrium evolutionary state and we quantify the amount of adaptive evolution in various classes of genomic sequence by their fitness flux. We continue by comparing this adaptive regime with the non-adaptive substitution dynamics under rapidly fluctuating selection, which has been a long-standing theme in population genetics. Finally, we contrast our model of adaptive evolution with Muller's ratchet, a classical model of ongoing deleterious evolution, and we discuss the causality between selection change and adaptive response in the presence of genetic drift.

Defining adaptation in molecular evolution

Here, we consider a phenotypic trait determined by several genomic loci, which is represented in Figure 1 by a continuous trait variable x. As a first case, we assume that selection on this trait is given by a time-independent single-peak fitness landscape $F(x)$ with optimal trait value x^*, as shown in Figure 1a. In this landscape, a mutation changing the trait value from x_1 to x_2 has the selection coefficient $\Delta F = F(x_2) - F(x_1)$. The trait value is assumed to be mostly fixed within a population (as in the example of the next section), but it changes over time by beneficial or deleterious substitutions at any of the contributing loci. This process is described by a path $x(t)$ as a function of evolutionary time t, shown as a line in Figure 1a. Because the fitness landscape is time-independent, evolution reaches an equilibrium state in which beneficial substitutions of a given selection coefficient occur at the same rate as deleterious substitutions of the opposite selection coefficient. In other words, the equilibrium distribution of selection coefficients for substitutions is symmetric, as shown in Figure 1b. In this state, beneficial changes merely

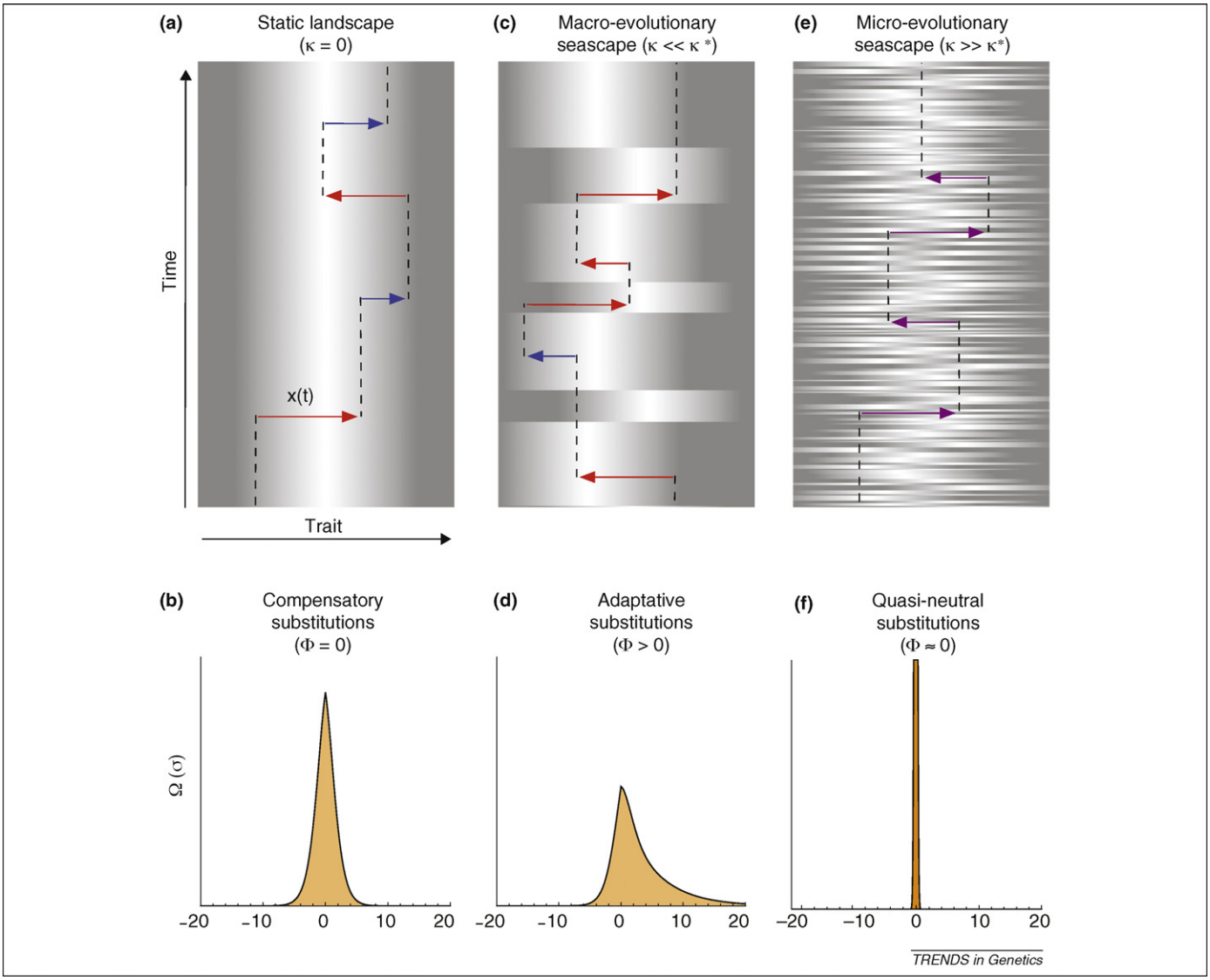

Figure 1. **(a,c,e)** Evolutionary path $x(t)$ of a quantitative trait. Selection on this trait is given by a fitness function $F(x,t)$ with fluctuation rate κ (lighter shading indicates fitter trait values). Beneficial substitutions in one of the genomic trait loci are marked by red arrows, deleterious substitutions by blue arrows and effectively neutral substitutions by violet arrows. **(b,d,f)** Distribution $\Omega(\sigma)$ of scaled selection coefficients $\sigma = 2N\Delta F$ of these substitutions. (a,b) Static fitness landscape ($\kappa = 0$): evolution reaches an equilibrium state in which beneficial changes merely compensate previous deleterious ones. The equilibrium distribution $\Omega(\sigma)$ is symmetric and the fitness flux Φ vanishes: there is no adaptation. (c,d) Macro-evolutionary fitness seascape ($\kappa << \kappa^*$): evolution reaches a non-equilibrium stationary state with a surplus of beneficial over deleterious substitutions. The distribution $\Omega(\sigma)$ is skewed toward positive values and there is a fitness flux $\Phi > 0$ quantifying adaptation. (e,f) Micro-evolutionary fitness seascape ($\kappa >> \kappa^*$): owing to multiple changes of selection during the fixation process, substitutions become quasi-neutral and are no longer adaptive.

Box 1. Equilibrium and non-equilibrium evolutionary dynamics

Here, we consider the evolution of a genomic locus at the population level. The sequence states of this locus are labeled as **a, b** etc. Each state **a** is assigned a fitness $F(\mathbf{a})$ first taken to be time-independent. Assuming that most of the population is fixed in a unique sequence state at most times, the macro-evolution of this system consists of substitutions between different sequence states. A substitution **a**→**b** takes place with a rate $u_{\mathbf{a}\to\mathbf{b}}$ that depends on its selection coefficient $\Delta F_{\mathbf{a}\to\mathbf{b}} = F(\mathbf{b}) - F(\mathbf{a})$; beneficial substitutions have higher rates than deleterious ones. If we consider an ensemble of loci evolving independently in the same sequence space and the same fitness landscape, we can define the probability that a locus is in a given sequence state, $Q(\mathbf{a})$. We now focus on pairs of sequence states **a**, **b** linked by a mutational opportunity. For any such pair, the ensemble-average rate of 'forward' substitutions **a**→**b** is given by $j_{\mathbf{a}\to\mathbf{b}} = Q(\mathbf{a})\ u_{\mathbf{a}\to\mathbf{b}}$, and the corresponding rate of 'backward' substitutions **b**→**a** is $j_{\mathbf{b}\to\mathbf{a}} = Q(\mathbf{b})\ u_{\mathbf{b}\to\mathbf{a}}$. We can then define the fitness flux between these sequence states as the product of selection coefficient and net mutation flux [16], $\Phi_{\mathbf{ab}} = \Delta F_{\mathbf{a}\to\mathbf{b}}\,(j_{\mathbf{a}\to\mathbf{b}} - j_{\mathbf{b}\to\mathbf{a}})$.

Equilibrium state

This state is a probability distribution Q_{eq} on sequence space satisfying the condition of detailed balance, which says that the forward substitution rate $j_{\mathbf{a}\to\mathbf{b}}$ and the backward rate $j_{\mathbf{b}\to\mathbf{a}}$ are equal:

$$Q_{eq}(\mathbf{a})u_{a\to b} = Q_{eq}(\mathbf{b})u_{b\to a} \quad \text{for any pair of sequence states } \mathbf{a},\mathbf{b}. \qquad \text{equation [1]}$$

This condition implies that the fitter of two sequence states (shown as lighter dot in Figure I) is always more likely than the less fit state: if $F(\mathbf{b}) > F(\mathbf{a})$, we have $Q_{eq}(\mathbf{b}) > Q_{eq}(\mathbf{a})$, exactly compensating the fact that the beneficial substitution rate is larger than the deleterious rate, $u_{\mathbf{a}\to\mathbf{b}} > u_{\mathbf{b}\to\mathbf{a}}$. The beneficial rate $j_{\mathbf{a}\to\mathbf{b}}$ is shown by the red arrow, the deleterious rate $j_{\mathbf{b}\to\mathbf{a}}$ by the blue arrow. Clearly, detailed balance also implies that Q_{eq} is time-independent and has vanishing fitness flux, $\Phi_{\mathbf{ab}} = 0$, between any pair of sequence states. This definition of equilibrium is well known in statistical physics, but it is more restrictive than the usage in most of the population genetics literature. The definition says that beneficial substitutions of selection coefficient ΔF occur at the same ensemble-average rate as deleterious substitutions of the opposite selection coefficient $-\Delta F$ and excludes the popular picture of many slightly deleterious substitutions balancing few strongly beneficial ones. It is instructive to compare the detailed balance of substitutions (i.e. fixed changes) with mutations in individual genomes. A mutation **a**→**b** occurs with a rate $\mu_{\mathbf{a}\to\mathbf{b}}$, which is independent of its fitness effect ΔF and, according to Kimura's classical result, equals the substitution rate under neutral evolution. The ensemble-average forward and backward mutation rates are $\nu_{\mathbf{a}\to\mathbf{b}} = Q_{eq}(\mathbf{a})\mu_{\mathbf{a}\to\mathbf{b}}$ and $\nu_{\mathbf{b}\to\mathbf{a}} = Q_{eq}(\mathbf{b})\mu_{\mathbf{b}\to\mathbf{a}}$. Hence, deleterious mutations are more frequent than beneficial ones: if $F(\mathbf{b}) > F(\mathbf{a})$, we have $\nu_{\mathbf{b}\to\mathbf{a}} > \nu_{\mathbf{a}\to\mathbf{b}}$ because $Q_{eq}(\mathbf{b}) > Q_{eq}(\mathbf{a})$ (this is obvious in the simplest case $\mu_{\mathbf{a}\to\mathbf{b}} = \mu_{\mathbf{b}\to\mathbf{a}}$ but holds more generally as long as variations in μ are uncorrelated with selection).

Non-equilibrium stationary state

We now consider evolution under a time-dependent fitness function $F(\mathbf{a}, t)$, which leads to time-dependent selection coefficients $F_{\mathbf{a}\to\mathbf{b}}(t)$ and substitution rates $u_{\mathbf{a}\to\mathbf{b}}(t)$. Figure II shows the minimal fitness seascape, where the selection coefficient $F_{\mathbf{a}\to\mathbf{b}}(t)$ changes sign but remains constant in magnitude. As long as the fitness function changes on macro-evolutionary time scales (i.e. remains constant during a fixation process in most cases), we can easily generalize the definition of the fitness flux to the form $\Phi_{\mathbf{ab}}(t) = \Delta F_{\mathbf{a}\to\mathbf{b}}(t)(j_{\mathbf{a}\to\mathbf{b}}(t) - j_{\mathbf{b}\to\mathbf{a}}(t))$. This definition can be generalized further to fitness changes on arbitrary time scales, such that every substitution **a**→**b** enters with the average fitness effect over the time interval (t_i, t_f) of its fixation process $\Delta F_{ave} = \int \Delta F_{\mathbf{a}\to\mathbf{b}}(t)dt/(t_f - t_i)$. If the time-dependence of the fitness function is defined by a stationary stochastic processes (as in the example discussed in this article), the ensemble of loci can still reach a stationary state with a time-independent probability distribution Q_s, but this state is no longer an equilibrium as defined above. Indeed, it can be shown that any stationary state under time-dependent selection is less adapted than the equilibrium state and has a positive total fitness flux, $\Phi \equiv \sum_{\mathbf{a},\mathbf{b}}\Phi_{\mathbf{ab}} > 0$ (i.e. there is a surplus of beneficial over deleterious changes). Intuition for this effect and for the different dynamical regimes of fitness seascapes can be gained from the example of Figure 1 in the main text.

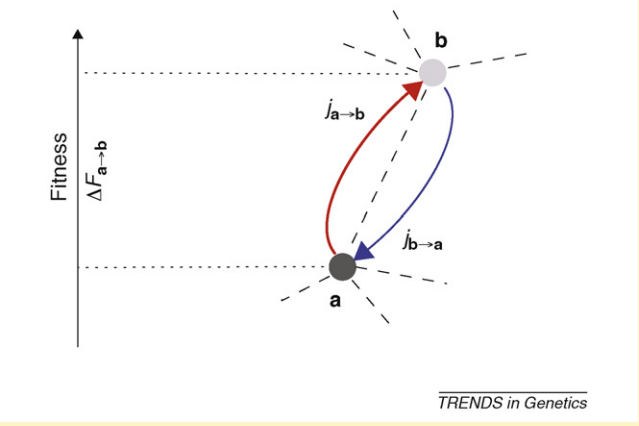

Figure I. Equilibrium state under time-independent selection.

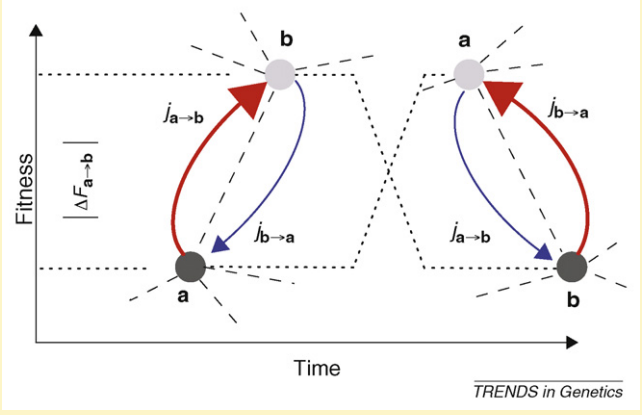

Figure II. Non-equilibrium steady state under time-dependent selection.

compensate previous deleterious ones, and the fitness remains constant on average: we can associate evolutionary equilibrium with the absence of adaptation. As illustrated in Box 1, the equilibrium state is defined by this so-called detailed balance between substitutions. We emphasize that detailed balance refers to fixed mutations and, hence, does not contradict the familiar notion that most mutations in individuals are deleterious. This property is, in fact, a consequence of equilibrium: because fitter sequence states are more likely to occur at equilibrium than less fit states, a random mutation is more likely to have a negative fitness effect than a positive. Hence, the distribution of selection coefficients for these mutations is skewed towards negative values. In our example of regulatory elements in yeast, we evaluate the distribution of selection coefficients both for mutations and substitutions from genomic data (Figure 2c). In Box 2, we show that an evolutionary dynamics of mutations, genetic drift and time-independent selection results in evolutionary equilibrium as a generic long-term outcome and we quantify

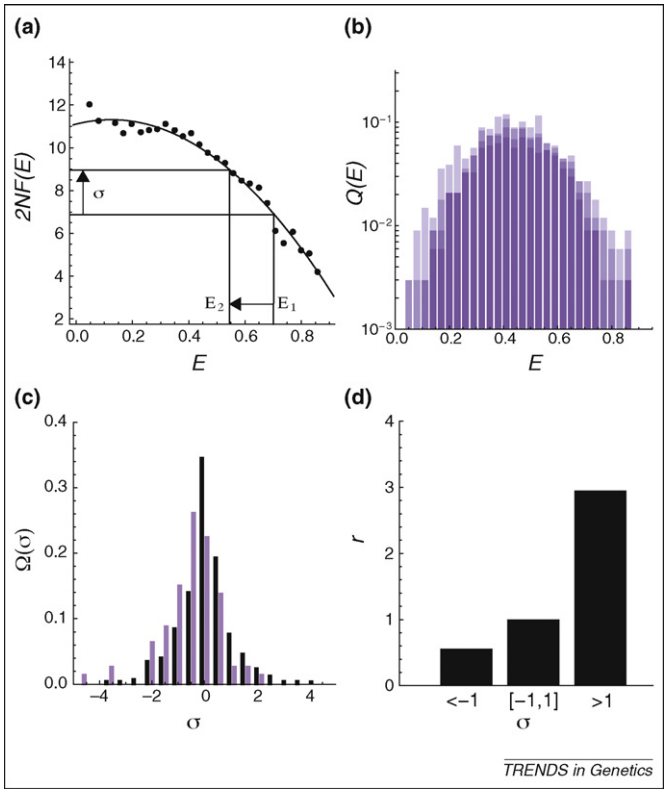

Figure 2. Evolution of Abf1-binding sites in yeast. **(a)** The scaled fitness landscape $F(E)$ as a function of binding energy E is inferred from energy distributions of functional sites and of background sequence in *S. cerevisiae* (binned data, dots; fit function, line; fitness values are scaled in units of the inverse effective population size $1/2N$) [34]. This landscape determines the scaled selection coefficient $\sigma = 2NF(E_2) - 2NF(E_1)$ of any mutation within binding sites as a function of the energies E_1 and E_2 of its sequence states. **(b)** The Abf1 site energy distribution $Q(E)$ is well conserved in four yeast sensu stricto species (superimposed purple bars, i.e. the parts with darkest shading are common to all species). **(c)** Distribution of cross-species fitness differences σ between pairs of orthologous sites with energy E_1 in *S. cerevisiae* and energy E_2 in *S. paradoxus* (black bars). The distribution is nearly symmetric, which is consistent with detailed balance between forward and backward substitutions (Box 1). Distribution of selection coefficients σ for polymorphisms in *S. cerevisiae* with ancestral allele E_1 and derived allele E_2 (purple bars), which are determined using *S. paradoxus* as an outgroup. The distribution is skewed towards negative values of σ because most mutations in individuals are slightly deleterious (Box 1). All fitness differences σ are evaluated using the fitness landscape of (a). **(d)** McDonald-Kreitman ratio $r = [(d/p)]/[d_0/p_0]$, where d is the number of point substitutions between *S. cerevisiae* and *S. paradoxus* and p is the number of *S. cerevisiae* polymorphisms in a given range of selection coefficients, $\sigma < -1$ (deleterious changes), $-1 < \sigma < 1$ (near-neutral changes defining d_0 and p_0) and $\sigma > 1$ (beneficial changes). The selection coefficient of a point substitution of energy change ΔE is evaluated as its average fitness effect in the ensemble of functional site sequences, $\sigma = \langle 2NF(E + \Delta E) - 2NF(E) \rangle$, and the selection coefficient of a polymorphism with ancestral allele E_1 and derived allele E_2 is defined as in (c). Polymorphism data are from the Saccharomyces Genome Resequencing Project [38].

the prevalence of moderately deleterious mutations over beneficial ones in this state.

Now, consider the same phenotypic trait evolving in a macro-evolutionary fitness seascape $F(x,t)$ (i.e. a fitness function that remains constant over the duration of most single substitution processes but changes on larger timescales, as shown in Figure 1c). This time-dependence of selection is in tune with our organismic picture of selection changes causing adaptive responses. The trait evolution of a population is again described by a path $x(t)$; individual substitutions are beneficial or deleterious with respect to the fitness function at the time of the substitution. However, the time-dependence of selection breaks the balance between beneficial and deleterious substitutions. After a change in selection, the trait value is likely to be less adapted than before and this can result in an adaptive substitution towards the new optimal value $x^*(t)$. Under on-going changes of selection, evolution can never reach equilibrium as defined by the detailed balance between beneficial and deleterious substitutions. Instead, the evolutionary process reaches a non-equilibrium stationary state marked by a surplus of beneficial over deleterious substitutions. The resulting distribution of selection coefficients for substitutions is biased towards positive values, as shown in Figure 1d. The surplus of beneficial over deleterious substitutions does not imply any increase in the mean fitness of a population over time. Rather, it is the result of changes in selection: this is the genomic signature of adaptive evolution. To measure the amount of adaptation per unit of time, we can define the fitness flux Φ as the product of total rate and average selection coefficient of substitutions. This quantity is always positive in a non-equilibrium stationary state and vanishes at equilibrium (Box 1). Positive values of Φ are inferred from sequence data in our example of the *Drosophila* genome.

Box 2. Mutation-selection-drift equilibria

Here, we provide a more quantitative introduction to evolutionary equilibria. We show that sequence evolution processes involving genetic drift (with effective population size N), mutations (with neutral rates $\mu_{\mathbf{a}\to\mathbf{b}}$ of a common order μ) and selection given by a time-independent fitness landscape $F(\mathbf{a})$ reach equilibrium under generic conditions, and we derive an explicit solution of the equilibrium probability distribution Q_{eq} on sequence space. We assume $\mu N << 1$ to be sufficiently small, so that macro-evolution consists of substitutions $\mathbf{a}\to\mathbf{b}$ between fixed sequence states, which is a reasonable assumption for many genomic evolution processes. According to the standard Kimura-Ohta theory, these substitutions take place with rates $u_{\mathbf{a}\to\mathbf{b}} = \mu_{\mathbf{a}\to\mathbf{b}}\,\phi(\Delta F_{\mathbf{a}\to\mathbf{b}}, N)$, where $\phi(\Delta F_{\mathbf{a}\to\mathbf{b}}, N) = 2N\Delta F_{\mathbf{a}\to\mathbf{b}}/(1 - e^{-2N\Delta F_{\mathbf{a}\to\mathbf{b}}})$. The substitution rate is enhanced with respect to the neutral rate for beneficial changes ($\Delta F_{\mathbf{a}\to\mathbf{b}} > 0$) and reduced for deleterious changes ($\Delta F_{\mathbf{a}\to\mathbf{b}} < 0$). Quite remarkably, the existence of an equilibrium state for the evolution process under selection depends only on the rates $\mu_{\mathbf{a}\to\mathbf{b}}$ of the corresponding neutral process, but not on the fitness landscape $F(\mathbf{a})$. If the neutral process has an equilibrium state Q^0_{eq} (for example, Q^0_{eq} = const. if all rates $\mu_{\mathbf{a}\to\mathbf{b}}$ are equal), an equilibrium also exists for evolution under an arbitrary time-independent fitness landscape $F(\mathbf{a})$. The equilibrium under selection takes the simple form [31,33]:

$$Q_{eq}(\mathbf{a}) = Q^0(\mathbf{a})e^{2NF(\mathbf{a})} \qquad \text{equation [2]}$$

This generalizes well-known single-site equilibrium distributions; for example, see Ref. [35]. Indeed, it is easy to see that the distribution Q_{eq} satisfies the detailed balance condition under selection (equation [1]), given that the distribution Q^0_{eq} satisfies the analogous condition under neutrality, $Q^0_{eq}(\mathbf{a})\,\mu_{\mathbf{a}\to\mathbf{b}} = Q^0_{eq}(\mathbf{b})\,\mu_{\mathbf{b}\to\mathbf{a}}$ for any pair of sequence states $\mathbf{a}$, $\mathbf{b}$. Selection enhances the occupation probability of the fitter state relative to the less fit state by a factor $[Q_{eq}(\mathbf{b})/Q_{eq}(\mathbf{a})]/[Q^0_{eq}(\mathbf{b})/Q^0_{eq}(\mathbf{a})] = e^{2N\Delta F_{\mathbf{a}\to\mathbf{b}}}$. This factor just matches the selection dependence of the ratio of Kimura-Ohta substitution rates, $\phi(\Delta F_{\mathbf{a}\to\mathbf{b}}, N)/\phi(\Delta F_{\mathbf{b}\to\mathbf{a}}, N) = e^{2N\Delta F_{\mathbf{a}\to\mathbf{b}}}$, leading to detailed balance of substitutions under selection. The inverse of this factor enters the ensemble-average mutation rates in individuals: the forward rate $\nu_{\mathbf{a}\to\mathbf{b}}$ is related to the backward rate $\nu_{\mathbf{b}\to\mathbf{a}}$ by $\nu_{\mathbf{a}\to\mathbf{b}}/\mu_{\mathbf{a}\to\mathbf{b}} = (\nu_{\mathbf{b}\to\mathbf{a}}/\mu_{\mathbf{b}\to\mathbf{a}})\,e^{-2N\Delta F_{\mathbf{a}\to\mathbf{b}}}$. Averaging over neutral rates, we obtain a relation between the total ensemble average rate $\nu(\Delta F)$ of beneficial mutations with selection coefficient $\Delta F > 0$ and the rate of deleterious mutations with opposite selection coefficient:

$$\nu(\Delta F) = \nu(-\Delta F)e^{-2N\Delta F} \qquad \text{equation [3]}$$

which quantifies the prevalence of moderately deleterious mutations over beneficial ones.

Not every time-dependent fitness function, however, causes adaptive substitutions. Figure 1e shows the evolution of the phenotypic trait in a micro-evolutionary fitness seascape $F(x,t)$, where fluctuations occur during single substitution processes. In this dynamics, substitutions become uncorrelated with the rapidly changing fitness optimum $x^*(t)$: they are no longer adaptive and become quasi-neutral (Figure 1f), in a sense to be made precise later.

Positive and negative selection at equilibrium: regulatory elements in yeast

Binding sites consist of approximately 10–15 contiguous nucleotides located upstream of protein-coding genes. If a transcription factor protein is bound to a functional site, it can regulate the downstream gene (i.e. enhance or repress its transcription). The interaction between factors and sites is well understood [24]; the probability of a factor molecule being bound to a site depends on the density of these molecules and on the binding energy E of the site, which in turn is determined by the site sequence. In other words, the binding energy E is a phenotypic trait, which characterizes the functionality of the site. The sequence dependence of E (i.e. the mapping from genotype to phenotype) can be measured directly [25,26] or can be inferred from sequences of functional sites [24,27] or promoter-binding assay experiments [28,29].

If we assume the evolution of binding sites is at equilibrium, we can go a step further and infer selection acting on this phenotype in the form of a fitness landscape $F(E)$ [30–32]. The genomic analysis underlying this inference is quite straightforward: for each intergenic subsequence with length equal to the binding-site length, we evaluate the binding energy E using its known sequence dependence. Hence, we obtain the normalized binding energy distribution $Q(E)$ for an ensemble of functional genomic sites, which are known binding sites or putative sites identified by evolutionary conservation, and the corresponding distribution $Q_0(E)$ for background sequence. At equilibrium, these distributions determine the average fitness of a site as a function of its binding energy, $F(E) = \log Q(E)/Q_0(E)$, by a simple generalization of equation [2] from sequence to phenotype distributions [31,32,33] (Box 2). For Abf1, a transcription factor with many target genes, the resulting fitness landscape of binding sites in the yeast *Saccharomyces cerevisiae* is shown in Figure 2a [34]. It determines the fitness effects of genomic changes in binding sites: a mutation inducing a shift of energy $\Delta E = E_2 - E_1$ has an average selection coefficient $\Delta F = F(E_2) - F(E_1)$. Because the fitness landscape is a monotonically decreasing function of the phenotype E, any mutation that reduces binding ($\Delta E > 0$) is, on average, deleterious ($\Delta F < 0$) and any mutation increasing binding is, on average, beneficial. This average fitness landscape is a minimal model for selection on binding sites, which attributes variation between different sites in one species and divergence between orthologous sites across species to genetic drift rather than to changes in selection. The fitness landscape assigns only moderate selection coefficients to individual substitutions, thereby enabling site sequences to evolve by deleterious changes in compensatory balance with beneficial ones. A conceptually similar inference of selection on individual point substitutions, rather than phenotype changes, has been performed in Ref. [35] for protein-coding sequence and Ref. [36] for binding sites. This analysis involves the further averaging over all sequence contexts of a given point substitution within binding sites.

The fitness landscape $F(E)$ provides a quantitative model for the evolution of binding sites. Its predictions and, hence, the assumption of evolutionary equilibrium underlying its inference can be tested against cross-species observations. For our example of Abf1-binding sites in yeast, we compare sequence and energy phenotype of orthologous sites in four *Saccharomyces* sensu stricto species [34]. In accordance with the moderate-selection landscape $F(E)$, there is considerable divergence of sequence and phenotype: on average, orthologous sites between the two most distant species *S. cerevisiae* and *S. bayanus* differ by approximately 3.5 point mutations and have energy differences of magnitude $|\Delta E| = 0.14$ (measured in units of the total energy range of functional sites) [34]. Nevertheless, the energy distributions of conserved functional sites shown in Figure 2b and the resulting fitness landscapes remain remarkably similar in all four species. Moreover, the distribution of cross-species fitness changes between orthologous sites in any two species is symmetric (i.e. there are equal numbers of beneficial energy changes with selection coefficient $\Delta F > 0$ and of deleterious changes with selection coefficient $-\Delta F$; Figure 2c). Both observations are consistent with the detailed balance condition of equilibrium and vanishing fitness flux, $\Phi = 0$ (Box 1). This balance includes binding-site turnover (i.e. the occasional loss and gain of functional sites) [34,37]. Recently available whole-genome polymorphism data in yeast [38] offer the unique possibility to relate the population genetics of these binding sites to their biophysical phenotype. Specifically, we evaluate single-nucleotide polymorphisms within *S. cerevisiae* Abf1 sites and estimate the selection coefficient ΔF of their point mutation from the energy-dependent fitness landscape $F(E)$ using the outgroup species *S. paradoxus* to determine the ancestral and the new (derived) allele. As expected, most of these polymorphisms have low population frequency because their derived allele arises from a recent mutation in an individual. The distribution of selection coefficients for polymorphisms is skewed towards negative values of ΔF (Figure 2c). This shows that most binding-site mutations in individuals are slightly deleterious, which is in accordance with the detailed balance of substitutions as discussed in Box 1. Furthermore, we can perform a McDonald-Kreitman analysis of Abf1-binding-site sequences, which infers selection from ratios of polymorphism and substitution numbers [7], and compare the results with the predictions obtained from the fitness landscape. This analysis confirms that mutations reducing binding ($\Delta E > 0$, hence $\Delta F < 0$) are predominantly under moderate negative selection and mutations that increase binding are under moderate positive selection, as shown in Figure 2d.

These observations provide evidence of multiple mutations under positive and negative selection in the

evolution of transcription-factor-binding sites. Yet, they are consistent with equilibrium in a time-independent fitness landscape, in which all positively selected mutations are compensatory and there is no adaptation. We emphasize again that this equilibrium is defined by phenotype, not sequence: compensations between deleterious and beneficial substitutions occur, in general, at different positions of a binding site. This kind of compensation is characteristic of quantitative traits at equilibrium (Figure 1a). Adaptive macro-evolution of binding sites would involve energy changes of the ensemble of binding sites, that is, changes in the distribution $Q(E)$ between species or of individual sites, which might leave the ensemble $Q(E)$ invariant, in response to a macro-evolutionary fitness seascape $F(E,t)$. Our present analysis does not exclude that some or even the majority of the observed cross-species changes in binding sites (including loss and gain of sites [34,37]) are adaptive in this sense. Our point is that positive selection alone does not yet prove adaptation – and it should not be over-interpreted as such.

Beyond equilibrium: dynamic selection and adaptation in fly genomes

In the previous example, the inferred selection was a static fitness landscape. It should be remembered, however, that the time-independence of selection was an assumption. This assumption leads to equilibrium as a minimal model that coherently captures positive and negative selection and explains the observed cross-species divergence. The reality of evolution is probably different in many cases. Changes in selection coefficients at a given genomic locus occur for various reasons, for example, changing external conditions or changes at another locus linked by epistasis. Here, we focus on selection with stochastic changes on macro-evolutionary time scales as a minimal model for adaptive evolution [16].

In a recent study of polymorphisms and substitutions in *Drosophila* genomes, we have found a surplus of beneficial over deleterious substitutions in both coding and non-coding regions of the genome, resulting in positive fitness flux values. Hence, these data show evidence of adaptive evolution and are incompatible with evolutionary equilibrium in any static fitness landscape [16]. In a minimal fitness seascape model, point mutations at individual genomic positions have selection coefficients $\Delta F(t)$ that randomly and independently change sign, according to a stochastic process with rate κ. This model stipulates that the strength of positive selection on a given genomic locus after a change in the fitness seascape is similar to the strength of negative selection before the change. The model is self-consistent: after an adaptive response, the strength of negative selection is again the same as the strength of positive selection before. For macro-evolutionary rates, which are much smaller than the average inverse polymorphism lifetime κ^*, the non-equilibrium stationary state of this model affords analytic solutions for the joint statistics of polymorphisms and substitutions, which we use for a Bayesian inference of the average strength and fluctuation rate of selection from genomic data. The qualitative features of this inference are quite intuitive. The polymorphism-substitution spectrum is a superposition of contributions from changes under negative and positive selection and near-neutral changes; the relative weight of these contributions is determined by strength and rate of selection. Evolutionary equilibrium ($\kappa = 0$) under selection leads to reduced counts of intermediate- and high-frequency polymorphisms and of substitutions compared with neutral evolution. A non-equilibrium state ($\kappa > 0$) has more changes under positive selection signaled by higher counts of high-frequency polymorphisms and substitutions than equilibrium at the same selection strength; the relative weight of positive-selection changes increases with the driving frequency κ. From this analysis, we infer that the *Drosophila* genome evolves under selection of substantial average strength with fluctuation rates κ comparable to the neutral point mutation rate (Figure 3). Time-dependent selection explains the observed fitness flux in a parsimonious way: selection changes open windows of positive selection, which trigger adaptive substitutions. However, the genome-wide analysis in *Drosophila* is constrained to individual genomic loci, unlike the phenotype-based inference in the previous section. A selection change at a given locus can be caused by an external change or by a substitution at another genomic locus coupled by fitness interactions. As an illustration of the second cause, consider again the evolution of a quantitative trait shown in Figure 1. A mutation that increases the trait value x is beneficial as long as $x(t) < x^*(t)$ (for example, before the first substitution in Figure 1a) but becomes deleterious when

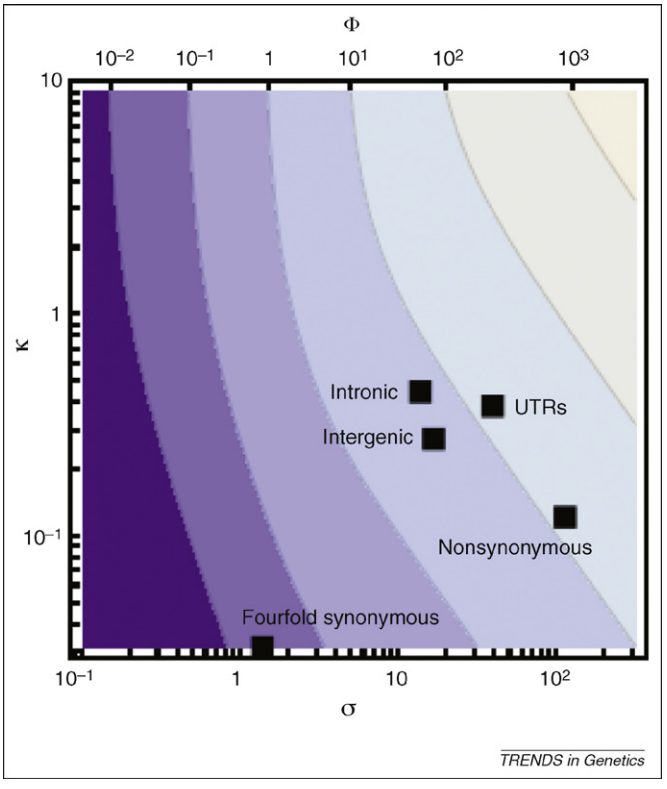

Figure 3. Adaptive evolution in the fly genome. The minimal fitness seascape $F(t) = \pm\sigma$ inferred for different sequence classes of *Drosophila melanogaster* (dots) has two parameters, the (average) amplitude σ (in units of $1/2N$) and the fluctuation rate κ (in units of μ) of selection coefficients. The fitness flux $\Phi(\sigma,\kappa)$ in the stationary state of this fitness seascape (in units of $\mu/2N$) is indicated by shading. All sequence classes of the *Drosophila* genome except fourfold synonymous sites are seen to evolve under substantial selection far from equilibrium with flux values $\Phi > 1$, providing quantitative evidence for adaptive evolution [16].

$x(t) > x^*(t)$ (after the first substitution in Figure 1a). Thus, compensatory evolution across loci can appear as adaptation if the coupling between these loci is unknown. The inference of time-dependent selection rests on the assumption that the positive fitness flux Φ is stationary (i.e. maintained over long evolutionary times). Under any time-independent selection model, positive values of Φ can only be transient and are not a generic feature of molecular evolution. The assumption of stationarity can be probed once polymorphism data become available for more species.

Several studies have identified positive selection from the statistics of polymorphisms and substitutions in *Drosophila* [7–9,12,15,19]. The analysis presented here goes a step further in quantifying adaptive evolution by its fitness flux Φ, which is determined by strength and fluctuation rate of a fitness seascape. Comparable values of Φ are inferred for different sequence classes of the *Drosophila melanogaster* genome, but the underlying selection parameters vary: the average strength of selection is greater in protein-coding DNA, whereas the fluctuation rate is greater in non-coding DNA [16] (Figure 3). The higher overall levels of selection compared with recent studies using the infinite-sites model (for example, see Refs [10,12]) are probably caused by the increased statistical power of our inference, which is based jointly on the full polymorphism spectrum, substitutions and sites conserved in the entire sample. Another recent study has inferred rate and strength of selective sweeps in the *Drosophila simulans* genome from an observed negative correlation between synonymous polymorphisms and non-synonymous substitutions [19]. The fitness flux Φ resulting from these sweeps (i.e. the product of selection strength and sweep rate) is of the same order of magnitude as the fitness flux inferred here [16]. The two methods address complementary regimes of selection strength and both are likely to underestimate the total fitness flux. A highly dynamic picture of the *Drosophila* genome arises: strong sweeps are the tip of the iceberg of adaptive evolution; at the same time, many adaptive substitutions occur at moderate selection coefficients. Of course, much more needs to be learned about the joint influence of epistatic interactions, genomic linkage and demographic history on the observed evolutionary pattern.

Micro-evolutionary fitness seascapes

Models of fluctuating selection have been a venerable subject of population genetics over the past decades [39–45]. Much of this work has been concerned with selection changes on micro-evolutionary time scales (i.e. with fluctuation rates κ larger than the average inverse polymorphism lifetime κ^*). Such time dependence of selection can arise from various ecological factors such as changing environment or lifestyle changes. It is discussed, for example, as a possible cause of the recent diversification of the human population since its migration out of Africa [46]. Another source of effective selection fluctuations is genetic draft: a polymorphic locus experiences temporary allele frequency shifts by linkage with neighboring loci under positive selection (hitchhiking), which are randomly interrupted by recombination [47]. The term 'fitness seascape' was initially coined for such micro-evolutionary fluctuations of selection [48].

The evolution of a quantitative trait in a micro-evolutionary fitness seascape is shown in Figure 1e. Substitutions cannot be adaptive in this dynamics because they do not respond to a coherent direction of selection during the fixation process. Clearly, this does not exclude adaptive response of allele frequencies to selection changes on micro-evolutionary scales. If we define the effective fitness effect ΔF of a substitution as the average of its time-dependent selection coefficient $\Delta F(t)$ during the fixation process, the distribution of ΔF becomes peaked at near-neutral values and the resulting fitness flux Φ becomes small [49] (Figure 1f). This averaging of time-dependent selection has been observed in population data of *Daphnia* and the scarlet tiger moth [50,51]. The quasi-neutrality of substitutions emerging in our micro-evolutionary seascape implies that forward and backward substitutions are equally likely at all times. However, the rate of substitutions is enhanced by selection fluctuations to values above the neutral rate [44] and attains a maximum if the fluctuation rate κ is of order κ^* [49]. This enhancement depends on population size and should be observable in a population bottleneck: ecological fluctuations, which are micro-evolutionary and, hence, well balanced in a large population, can cause a dramatic increase in the substitution rate during a bottleneck when the population size is temporarily reduced. A similar enhancement is induced if selection varies spatially and the number of environments occupied decreases during a bottleneck [41]. These increases in substitution rate go beyond the well-known effect of a bottleneck under time-independent selection, where the substitution rate can reach near-neutral values by temporary removal of selective constraint. Micro-evolutionary selection fluctuations also generate a non-neutral spectrum of polymorphism frequencies [45]. This, together with the increase in substitution rates, can confound population-genetic tests of positive selection [52].

Adaptation and Muller's ratchet

Here, we argue for a sharpened concept of adaptive evolution at the molecular level. Adaptation requires positive selection, but not every mutation under positive selection is adaptive. Selection and adaptation always refer to a molecular phenotype depending on a single genomic locus or on multiple loci, such as the energy of a transcription-factor-binding site in our first example. This correlates the direction of selection at all loci contributing to the phenotype and calls for the distinction between adaptation and compensation. The infinite-sites approximation, which is contained in many population-genetic models, neglects such correlations and is therefore not optimally suited to infer adaptation [16,23]. Here, we address this problem by a joint dynamical approach to selection and genomic response in a genome with finite number of sites. In this approach, adaptive evolution is characterized by a positive fitness flux Φ, which measures the surplus of beneficial over deleterious substitutions.

It is instructive to contrast this view of adaptive evolution with Muller's ratchet, a classical model of evolution by deleterious substitutions [53,54]. This model postulates

a well-adapted initial state of the genome so that all, or the vast majority of, mutations have negative fitness effects. Continuous fixations of slightly deleterious changes then lead to a stationary decline in fitness (i.e. to negative values of Φ). Similarly to the infinite-sites approximation, this model neglects compensatory mutations. In a picture of a finite number of sites, it becomes clear that every deleterious substitution leads to the opportunity for at least one compensatory beneficial mutation (or more, if the locus contributes to a quantitative trait), so that the rate of beneficial substitutions increases with decreasing fitness. Therefore, assuming selection is time-independent, decline of fitness ($\Phi < 0$) is only a transient state and the genome will eventually reach detailed balance between deleterious and beneficial substitutions, that is, evolutionary equilibrium ($\Phi = 0$). As long as selection is time-independent, an equilibrium state exists for freely recombining loci and in a strongly linked (i.e. weakly recombining) genome, although its form is altered in the latter case by interference selection [55,56]. Conversely, an initially poorly adapted system will have a transient state of adaptive evolution ($\Phi > 0$) before reaching equilibrium. Time-dependent selection, however, continuously opens new windows of positive selection, the genome is always less adapted than at equilibrium and the adaptive state becomes stationary. Thus, we reach a conclusion contrary to Muller's ratchet. Because selection in biological systems is generically time-dependent, decline of fitness is less likely even as a transient state than suggested by Muller's ratchet: the model offers no explanation of how a well-adapted initial state without opportunities of beneficial mutations is reached in the first place.

As a minimal model for adaptive evolution, we have introduced the Fisher-Wright process in a macro-evolutionary fitness seascape, which is defined by stochastic changes of selection coefficients at individual genomic positions on time scales larger than the fixation time of polymorphisms (and is thus different from micro-evolutionary selection fluctuations and genetic draft). Time-dependence of selection is required to maintain fitness flux: the seascape model is the simplest model that has a non-equilibrium stationary state with positive Φ. The two parameters of the minimal model (strength and rate of selection changes) are clearly just summary variables for a much more complex reality. The vastly larger genomic datasets within and across species will enable us to infer the dynamics of selection beyond this minimal model.

The arrow of time in molecular evolution

Our phenotypic picture of evolution implies causality between individual selection changes and the adaptive response to these changes. At the molecular level, this causality is maintained for genetic sweeps, which are the genomic response to strong new selection. At moderate levels of selection, causality is lost for individual genomic changes because selection competes with genetic drift and hitchhiking. However, there is still a statistical causality between selection change and genomic change, which is marked by a surplus of adaptive over deleterious substitutions as measured by the fitness flux Φ. Causality implies a temporal order between selection change and adaptive response, which defines the direction of evolutionary time. This becomes intuitively clear if we look at the substitution paths of Figure 1. Reverting the direction of time (that is, following these paths backwards from future to past) converts every deleterious substitution into a beneficial one and vice versa, which changes Φ into $-\Phi$. The equilibrium substitution path of the Fisher-Wright model under time-independent selection is converted into a statistically equivalent path because detailed balance implies equal numbers of beneficial and deleterious substitutions (i.e. $\Phi = -\Phi = 0$; Figure 1a). The same is true for the effectively neutral substitution dynamics in a micro-evolutionary fitness seascape shown in Figure 1e. The non-equilibrium path of Figure 1c, however, has $\Phi > 0$ and becomes meaningless in reverse order because 'adaptive' substitutions would occur before selection changes and would continuously lower the fitness ($-\Phi < 0$). Hence, time-dependent selection imprints the adaptive arrow of time on molecular evolution.

Acknowledgements

We enjoyed numerous valuable discussions on the topics of this article with participants of the workshop 'Population Genetics and Genomics' held in fall 2008 at the Kavli Institute for Theoretical Physics (KITP, UC Santa Barbara) and we thank KITP for its hospitality during that workshop. This work has been partially supported by DFG grant SFB 680, and by NSF grant PHY05-51164 (to KITP).

References

1 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press
2 Gillespie, J.H. (1991) *The Causes of Molecular Evolution.* Oxford University Press
3 Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286
4 Nei, M. (2005) Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* 22, 2318–2342
5 Li, W.H. (1997) *Molecular Evolution.* Sinauer Press
6 Haerty, W. *et al.* (2007) Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177, 1321–1335
7 McDonald, J. and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654
8 Smith, N.G. and Eyre-Walker, A. (2002) Adaptive protein evolution in *Drosophila*. *Nature* 415, 1022–1024
9 Fay, J.C. *et al.* (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415, 1024–1026
10 Sawyer, S.A. *et al.* (2003) Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J. Mol. Evol.* 57, S154–S164
11 Eyre-Walker, A. (2006) The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21, 569–575
12 Sawyer, S.A. *et al.* (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6504–6510
13 Larracuente, A.M. *et al.* (2008) Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24, 114–123
14 Kohn, M.H. *et al.* (2004) Inference of positive and negative selection on the 5′ regulatory regions of *Drosophila* genes. *Mol. Biol. Evol.* 21, 374–383
15 Andolfatto, P. (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437, 1149–1152
16 Mustonen, V. and Lässig, M. (2007) Adaptations to fluctuating selection in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2277–2282
17 Schlötterer, C. (2003) Hitchhiking mapping functional genomics from the population genetics perspective. *Trends Genet.* 19, 32–38
18 Glinka, S. *et al.* (2003) Demography and natural selection have shaped genetic variation in *Drosophila melanogaster* A multi-locus approach. *Genetics* 165, 1269–1278

19 Macpherson, J.M. *et al.* (2007) Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophila*. *Genetics* 177, 2083–2099
20 Teschke, M. *et al.* (2008) Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. *Genetics* 180, 1537–1545
21 Crow, J.F. and Kimura, M. (1970) *An Introduction to Population Genetics Theory*. Harper & Row
22 Sawyer, S.A. and Hartl, D.L. (1992) Population genetics of polymorphism and divergence. *Genetics* 132, 1161–1176
23 Desai, M.M. and Plotkin, J.B. (2008) The polymorphism frequency spectrum of finitely many sites under selection. *Genetics* 180, 2175–2191
24 Berg, O.G. and von Hippel, P.H. (1987) Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* 193, 723–750
25 Fields, D. *et al.* (1997) Quantitative specificity of the Mnt repressor. *J. Mol. Biol.* 271, 178–194
26 Maerkl, S.J. and Quake, S.R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315, 233–237
27 Stormo, G.D. and Fields, D. (1998) Specificity, free energy and information content in protein DNA interactions. *Trends Biochem. Sci.* 23, 109–113
28 Kinney, J.B. *et al.* (2006) Precise physical models of protein-DNA interaction from high-throughput data. *Proc. Natl. Acad. Sci. U. S. A.* 104, 501–506
29 Foat, B.C. *et al.* (2006) Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* 22, 141–149
30 Gerland, U. and Hwa, T. (2002) On the selection and evolution of regulatory DNA motifs. *J. Mol. Evol.* 55, 386–400
31 Berg, J. *et al.* (2004) Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* 4, 42
32 Mustonen, V. and Lässig, M. (2005) Evolutionary population genetics of promoters: Predicting binding sites and functional phylogenies. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15936–15941
33 Berg, J. and Lässig, M. (2003) Stochastic evolution and transcription factor binding sites. *Biophysics (Oxf.)* 48 (Suppl. 1), 36–44
34 Mustonen, V. *et al.* (2008) Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2277–2282
35 Halpern, A.L. and Bruno, W.J. (1998) Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol. Biol. Evol.* 15, 910–917
36 Moses, A.M. *et al.* (2003) Position specific variation in the rate of evolution in transcription factor binding sites. *BMC Evol. Biol.* 3, 19
37 Doniger, S.W. and Fay, J.C. (2007) Frequent gain and loss of functional transcription factor binding sites. *PLOS Comput. Biol.* 3, e99
38 Carter, D. *et al.* (2008) Population genomics of domestic and wild yeasts. *Nat. Precedings* http://hdl.handle.net/10101/npre.2008.1988.1
39 Wright, S. (1948) On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* 2, 279–294
40 Kimura, M. (1954) Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics* 39, 280–295
41 Ohta, T. (1972) Population size and rate of evolution. *J. Mol. Evol.* 1, 305–314
42 Gillespie, J.H. (1972) The effects of stochastic environments on allele frequencies in natural populations. *Theor. Popul. Biol.* 3, 241–248
43 Gillespie, J.H. (1993) Substitution processes in molecular evolution. *I. Uniform and clustered substitutions in a haploid model. Genetics* 134, 971–981
44 Takahata, N. *et al.* (1975) Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc. Natl. Acad. Sci. U. S. A.* 72, 4541–4545
45 Takahata, N. and Kimura, M. (1979) Genetic variability maintained in a finite population under mutation and autocorrelated random fluctuation of selection intensity. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5813–5817
46 Voight, B.F. *et al.* (2006) A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72
47 Gillespie, J.H. (2001) Is the population size of a species relevant to its evolution? *Evolution* 55, 2161–2169
48 Merrell, D.J. (1994) *The Adaptive Seascape*. Minnesota University Press
49 Mustonen, V. and Lässig, M. (2008) Molecular evolution under fitness fluctuations. *Phys. Rev. Lett.* 100, 108101
50 Lynch, M. (1987) The consequences of fluctuating selection for isozyme polymorphisms in *Daphnia*. *Genetics* 115, 657–669
51 O'Hara, R.B. (2005) Comparing the effects of genetic drift and fluctuating selection on genotype frequency changes in the scarlet tiger moth. *Proc. Biol. Sci.* 272, 211–217
52 Huerta-Sanchez, E. *et al.* (2008) Population genetics of polymorphism and divergence under fluctuating selection. *Genetics* 178, 325–337
53 Muller, H.J. (1932) Some genetic aspects of sex. *Am. Nat.* 66, 118–138
54 Felsenstein, J. (1974) The evolutionary advantage of recombination. *Genetics* 78, 737–756
55 Hill, W.G. and Robertson, A. (1966) The effect of linkage on the limits to artificial selection. *Genet. Res.* 8, 269–294
56 Comeron, J.M. and Kreitman, M. (2000) The correlation between intron length and recombination in *Drosophila*: dynamic equilibrium between mutational and selective forces. *Genetics* 156, 1175–1190