

The contribution of statistical physics to evolutionary biology

Harold P. de Vladar¹ and Nicholas H. Barton^{1,2}

¹I.S.T. Austria, A-3400 Klosterneuburg, Austria

²Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Evolutionary biology shares many concepts with statistical physics: both deal with populations, whether of molecules or organisms, and both seek to simplify evolution in very many dimensions. Often, methodologies have undergone parallel and independent development, as with stochastic methods in population genetics. Here, we discuss aspects of population genetics that have embraced methods from physics: non-equilibrium statistical mechanics, travelling waves and Monte-Carlo methods, among others, have been used to study polygenic evolution, rates of adaptation and range expansions. These applications indicate that evolutionary biology can further benefit from interactions with other areas of statistical physics; for example, by following the distribution of paths taken by a population through time.

Parallel foundations of evolution and statistical physics

During the late 19th century, Boltzmann established the theoretical foundations of statistical mechanics (see [Glossary](#)), in which the behaviour of ensembles of particles explains large-scale phenomena [1]. For example, the position and velocity of the particles in a gas can fluctuate between very many states (termed ‘microstates’), but averages over all the configurations that give the same observable macroscopic states (temperature and pressure, say) [2]. A similar averaging over equivalent microstates is made in both population and quantitative genetics: one averages over individual gene combinations to describe a population by its allele frequencies, and can further average over all the allele frequencies that are consistent with a given mean and variance of a quantitative trait. In this sense, physicists and evolutionary biologists both model populations (a gas or a gene pool) rather than precise types (individual particles or genotypes). This ‘statistical’ description in terms of a few variables, the macrostates, summarizes the many possible configurations of the microstates (degrees of freedom), which cannot be accurately measured or described. Furthermore, the macrostates are then sufficient to predict other properties without reference to the microstates. For example, thermodynamics describes macroscopic properties without referring to individual particles; similarly, quantitative genetics does not refer to allele frequencies to predict the trait mean in the next generation.

Hence, evolutionary biology and statistical physics often use similar theoretical methodologies, although studying

Glossary

Boltzmann distribution: a probability measure of the microscopic states of a physical system that is composed of classical (i.e. not quantum) particles in thermodynamic equilibrium. This distribution has a density proportional to the factor $\exp(-E/kT)$, where E is the energy of a state, k is Boltzmann’s constant and T is the absolute temperature.

Detailed balance: an equilibrium where the probability flux of the transitions between any two states is equal in either direction. In population genetics, this implies that the numbers of adaptive and deleterious substitutions have to be equal on average.

Entropy: a measure of the number of possible configurations of a system. The classical measure of entropy is due to Boltzmann: $S = -k \log \Omega$, where Ω is the number (or density) of microscopic states (e.g. allele frequencies) that a system can realize for a given macroscopic state (mean fitness, a quantitative variable, etc.) and k is Boltzmann’s constant. Relative entropy is defined as $S = -\int \psi \log(\psi/\varphi) d\mathbf{p}$, where the \mathbf{p} are the microscopic states, and the sum goes over all possible realizations; ψ is the distribution of microstates, and φ is a base or reference distribution (satisfying $\varphi = 2NV_{sp}$). However, when $\varphi = \text{const.}$ we have Shannon’s entropy, which is the form used in statistical physics. Entropy is also equivalent to the log-likelihood of φ (the proposed distribution), and ψ is the sampling probability of the actual distribution.

Fisher’s information: a measure of how much an infinitesimal change in an unknown parameter θ affects the likelihood ψ of an observed data set, \mathbf{p} . Fisher’s information is defined as $F = \int \psi(\mathbf{p}; \theta) \left(\frac{\partial}{\partial \theta} \log[\psi(\mathbf{p}; \theta)] \right)^2 d\mathbf{p}$. When the parameter θ is time, Fisher’s information describes the amount of information gained through selection.

Fitness flux: a measure of adaptation defined as $\phi(t) = s(p,t)dp/dt$, where s is the selection coefficient (fitness gradient) and p is the allelic frequency. Geometrically, it is the strength of fitness change (given that s is the gradient if mean fitness, \bar{W} , along the direction of evolution (given by dp/dt). The cumulative fitness flux, $\Phi = \int \phi dt$, is a measure of the total amount of adaptation through the history of a population.

Free fitness: the expected gain in log-mean fitness after selection; after an analogy with the free energy of a physical system, which is the amount of work that can be done in a thermodynamic system. Free fitness (I) emerges naturally when computing the gain in entropy S after an allele or a trait underwent selection [48], and has an equivalent expression to free energy, that is, $I = \langle \log(\bar{W}) \rangle - S/2N$ (in physics, $\langle \log(\bar{W}) \rangle$ should be replaced by $\langle E \rangle$, and $2N$ by $1/kT$; see entry for Boltzmann distribution).

Hill–Robertson interference: interference in the selective sweep of an allele, owing to the selective effects at another linked loci. Hill–Robertson interference implies that, in the presence of recombination, genotypes with multiple mutations arise more easily by recombining existing single mutations than by multiple mutation events.

Path ensemble: a formalism of nonequilibrium statistical mechanics and quantum mechanics where the description of the system emphasizes not the states of a population of entities, but the distribution of possible stochastic paths that such a population can follow.

Quasispecies: population of replicators (typically asexual) with a high genotypic variability maintained by elevated mutation rates.

Replicator dynamics: dynamical equations that describe the change in time the frequency p of the different types (in particular genotypes). It has the general form $dp/dt = p\Delta W + T$, where ΔW is the difference between the fitness of the type and the mean fitness, and T are the ‘transmission’ terms, that can involve mutation, migration, recombination, etc.

Selective death: failure to survive or reproduce owing to differences in genotype.

Stationary distribution: a probability distribution that does not change in time. This is found from the diffusion equation by setting $\partial\psi/\partial t = 0$ and solving the resulting differential equation that is independent of time. A stationary solution might not exist (e.g. if selection is changing in time in particular ways) and, if it exists, it might require detailed balance.

Corresponding author: de Vladar, H.P. (hpvladar@ist.ac.at)

Statistical mechanics: a mathematical framework explaining the relationship between the macroscopic properties of a system, in terms of the dynamics of the microscopic variables. At equilibrium, it leads to the classical concepts of entropy, free energy and temperature, for example. Out of equilibrium, these quantities cannot be defined formally, and current research focuses on finding probabilistic measures that apply in general, but are still based on the microscopic dynamics. Based principally on the properties of stochastic processes (e.g. the diffusion equations, or path ensembles), these measures can be applied to the distribution of allele frequencies (e.g. Fisher's information and fitness flux).

Travelling waves: solutions to nonlinear differential equations characterized by functions that are of stable shape, and move at a certain velocity either in physical space, or in genetic space. (Travelling waves are also known as solitons in the physics and mathematics literature.)

Truncation selection: Scheme where individuals that have traits outside a prescribed range are eliminated. This type of selection is popular in artificial selection.

very different phenomena. We argue here that there are close analogies between evolutionary genetics and statistical physics. Physical techniques had an early influence on molecular biology (Box 1). However, more specifically, nonequilibrium methods are based on the same theory of stochastic processes that is used in population genetics. Thus, some physical theories promise further developments that can deepen understanding of evolution in two ways: either by applying common mathematical tech-

Box 1. Evolution and the material basis of heredity

Early during the 20th century, Fisher embarked on the mathematical formalization of the Mendelian principles of heredity, following the earlier development of biometrics by Galton, Pearson and Weldon, all with the aim of quantifying evolution by natural selection. Fisher used the diffusion approximation (Box 2) to describe the evolution of allele frequencies [18]. In 1929, he introduced the Fundamental Theorem of Natural Selection [71]; comparing it to the second law of thermodynamics, the increase of entropy, he intended the theorem to be an exact result, a 'biological law' [81,82]. Although his comparison with the second law is flawed, it shows how the quantitative approach to heredity was influenced by statistical thermodynamics (indeed, Fisher had studied with E.T. Jeans, a physicist). That mechanistic basis of evolution, population genetics, was formulated without knowledge of the physical nature of the Mendelian genes, which was still unknown during the 1930s: the structure of DNA was not established until 1953.

In the following decades, Delbrück, formerly an astrophysicist, started a collaboration using ionizing radiation on *Drosophila* to understand the physical nature of the genes, as a working system to try to identify fundamental physical laws that would account for living and nonliving matter [83]. Later, the ingenious Luria-Delbrück experiment proved the basis of darwinian evolution. They performed a statistical comparison between the number of bacteria developing resistance to lysogenic viruses and its expected distribution, which was derived from a mathematical analysis, an unusual quantitative approach for the biologists of the time [84]. Soon after, the quantum theorist Schrödinger published *What is Life?* [85], posing fundamental biological questions in physicists' language, partly based on Delbrück's discoveries. This book gave strong motivation to the first molecular biologists (among them Perutz, Wilkins, Crick and Watson) to find how DNA transmitted the heritable information to future generations [86]. Molecular biology was influenced, in large part, by the use of physical techniques such as X-ray crystallography to determine biological structures. Evolution, however, while resting on that material basis of DNA, is not explained by it. Indeed, the population genetic framework that is used today was developed before the discovery of the structure of DNA, and was not changed by the establishment of molecular biology, because it rests only on Mendel's laws. However, the theoretical methods that are common to statistical physics and to evolutionary biology give a deeper understanding of the evolutionary consequences of heredity.

niques (e.g. diffusion equations; Box 2), or by developing precise analogies that incorporate new concepts (e.g. ensemble averaging, information or entropy). These techniques are being applied in different aspects of evolutionary biology. In this article, we focus mainly on those that we consider most promising for population genetics. We aim to introduce the reader to these methods by reviewing representative examples in the literature.

The cost of selection, entropy and information

It is extraordinary that the selection of random mutations has created complex organisms that appear exquisitely designed to fit their environment. Selection can be seen as taking information from the environment, and coding it into the DNA sequence [3]: thus, the gene pool contains information about those specific sequences that confer high fitness. This idea can be quantified using the concepts of entropy and information [4]. Entropy is a measure of the number of different states in which a population is likely to be found: thus, selection of one specific genotype, or genotype frequency, corresponds to minimal entropy. Asking

Box 2. The diffusion equation

The diffusion equation originated in Bachelier's models of fluctuations in share prices in 1900, and was rediscovered 73 years later as the Black-Scholes formula, disastrously popular among economists. Diffusion theory in economics is equivalent to the theory of Brownian motion, devised by Einstein to explain random molecular collisions, and soon after extended in physical applications [87]. Fisher [18] compared mendelian genetics to 'the theory of gases' and introduced the diffusion methods for the allele frequencies. Kolmogorov [88] gave a more formal approach to selection and drift. Kimura [89] later extended this formalism to nonequilibrium cases. For population genetics, the diffusion equation is a rather convenient representation of evolution of finite populations where genetic drift is present. We could choose to model the change in allele frequencies directly, in what is known as a Wright-Fisher process. However, genetic drift evolves stochastically, making the outcomes of evolution unpredictable. The diffusion equation gauges these outcomes in a probabilistic way, describing the distribution of allele frequencies at each time. (A third way to describe an evolving populations is to use the whole history as a random variable instead of the allele frequencies at a time, in what is known as a path ensemble; Box 3). In short, the diffusion equation is a partial differential equation describing the change in time of the probability density of the allele frequencies p , namely

$$\frac{\partial \psi}{\partial t} = -\frac{\partial}{\partial p} \left(M_{\delta p} \psi \right) + \frac{1}{2} \frac{\partial^2}{\partial p^2} \left(V_{\delta p} \psi \right),$$

where $M_{\delta p}$ are the deterministic factors, owing to selection, mutation, migration, and so on, and $V_{\delta p}$ is the variance of the fluctuations by drift usually of the form $p(1-p)/2N$. Making the left-hand side equal to zero leads to the stationary solution derived by Wright by other means [16]:

$$\psi = C \bar{W}^{2N} [p(1-p)]^{4N\mu-1}$$

For details of the derivation, see [89]. In particular, the term \bar{W} defines the 'fitness landscape', which can be thought of as a surface in the space of allele frequency (Figure 1, main text), or in the quantitative variables (Figure 2, main text).

The diffusion equation, the coalescent process and the path ensemble all describe the same process and are mathematically equivalent. Each has different advantages and limitations; whereas a stochastic differential equation, the diffusion equation and the path ensemble do not require detailed balance, the stationary distribution above, does. Yet, this solution is exact, quite general and relatively simple.

how the genotype of an individual, or a population, depends on the selection that they have experienced, can be quantified by an entropy that measures how strongly selection has clustered the population around a specific genotype. Haldane [5] showed that the number of selective deaths needed to fix an allele is independent of the selection pressure, and Kimura [6] pointed out that Haldane's 'cost of natural selection' is exactly the information gained by fixing a specific allele. This relation applies very generally to asexual populations [7] but fails with recombination (see below). The theory of quasispecies (a model of mutation–selection balance) emphasizes that the reproductive rate (selection) limits the amount of information that can be maintained in the face of random mutations [8]. However, this constraint can be relaxed with recombination and epistasis [9]. Analogies with statistical physics help one to understand how selection accumulates information. We first consider infinite populations (evolving deterministically) and then the more general case where random drift in finite populations drives evolution.

Deterministic evolution and the role of recombination

The information content of a single large population that is evolving deterministically is measured by the entropy, defined as $S = -\sum_x p_x \log(p_x)$, where p_x is the frequency of alleles or genotype x [5,8]. This entropy reflects the information accumulated and maintained by evolution, and is closely related to Shannon's information [3,4]. Remarkably, the replicator dynamics can be obtained by maximizing a different measure, (Fisher information: $F = \sum_x p_x \left(\frac{d}{dt} \log(p_x)\right)^2$), which measures divergence between two distributions [10,11]. Fisher's information is the 'acceleration' of the entropy (i.e. $d^2 S/dt^2 = F$), where $\frac{d}{dt} \log(p_x)$ is interpreted as the information contained about selection when we observe how the frequency of x has changed. For example, for a beneficial allele under selection, this would be proportional to the selective value s . Normally, we predict the change of the frequency distribution when we know s . Fisher's information takes the parent and the offspring distributions as given, and measures the effect of selection from the difference between these two [11].

Sexual versus asexual reproduction

It has long been understood that, when combined with truncation selection, sexual reproduction is more efficient than asexual in fixing beneficial genotypes [12,13]. In the former case, the maximum information increases as $n^{1/2}$ (n being the number of loci), whereas in the latter, it increases by only one unit per generation [3,14]. The maximum number of loci that can be maintained despite the randomizing effect of mutation is $1/\mu$ for asexuals [14], whereas for sexuals (with free recombination) it can be as high as $1/\mu^2$, where μ is the mutation rate at each locus [14,15]. (More loci would produce more mutants and, hence, decrease the amount of information.) According to Haldane's principle [16], every deleterious mutation must be eliminated by a failure to reproduce (a 'selective death'). Therefore, the mutation load is independent of selection strength, and is half as great if selection eliminates two copies in a recessive homozygote at the same time. In haploids, redundancy leads to a similar gain in efficiency [15,17].

Stochastic evolution: the diffusion of allele frequencies

The diffusion approximation shows how the distribution of allele frequencies at many loci changes through time in finite populations. In this case, selection, mutation and migration are modelled as deterministic factors, and genetic drift introduces random fluctuations to populations within an ensemble (Box 2). (Other treatments are possible, where mutations are regarded as carrying random changes to individuals within a single population.) In other fields, constant diffusion coefficients have been widely used, leading to simple Gaussian solutions (Box 2). However, Gaussian solutions are not appropriate for population genetics, because allele frequencies range between zero and one, and sometimes cluster near fixation, in a bimodal distribution. After its introduction by Fisher [18], Kolmogorov applied the more general diffusion method to the neutral island model [19], which Wright [20] had already solved by different means. Kimura relied on the diffusion approximation to model the evolution of finite populations [21] and for his neutral theory of molecular evolution [22].

The diffusion approximation, central to both population genetics and statistical physics, provides a way to model many factors in a mathematically tractable way. Crucially, it approximates a wide variety of more detailed models. Mathematically, it is equivalent to the coalescent process that describes the evolution of samples from a population, and to path ensemble methods that describe the distribution of population histories (see below). In physics, diffusion equations describe nonequilibrium processes and are hard to relate to quantities such as temperature, entropy, or free energy, which are well defined only in thermodynamic equilibrium through the Boltzmann distribution.

Wright showed that selection, mutation and drift give an explicit distribution, proportional to \bar{W}^{2N} , where \bar{W} is the mean fitness of a population of size N [20]. This is closely analogous to the Boltzmann distribution ($\sim e^{-E/kT}$) [1,2], with $\log(\bar{W})$ corresponding to (negative) energy, $-E$, and $1/2N$ to the temperature, kT (Box 2). This result was the basis for Wright's metaphor of an adaptive landscape: a surface of mean fitness laid over the multidimensional space of allele frequencies [23] (Box 2; Figure 1).

Jumps between adaptive peaks

When the stationary distribution is clustered around alternative peaks in the adaptive landscape, the rate at which random drift causes shifts between these states is approximated by a general formula that is proportional to the probability of being at the saddle point (adaptive valley) that separates them, and to the leading eigenvalue that describes the instability at that point [24,25]. Wright [25] worked out transition rates for chromosome rearrangements, ideas rigorously formulated later using diffusions [26]. Rouhani and Barton [24,27] found the rate of peak shifts in a spatially structured population, borrowing from an identical analysis of transitions between alternative vacuum states.

Travelling waves

The distribution of a quantitative trait, or of fitness itself, can be seen as a travelling wave that travels at a steady rate as the population adapts, either in actual or in phenotypic space. Most analyses have been of asexuals, which increase

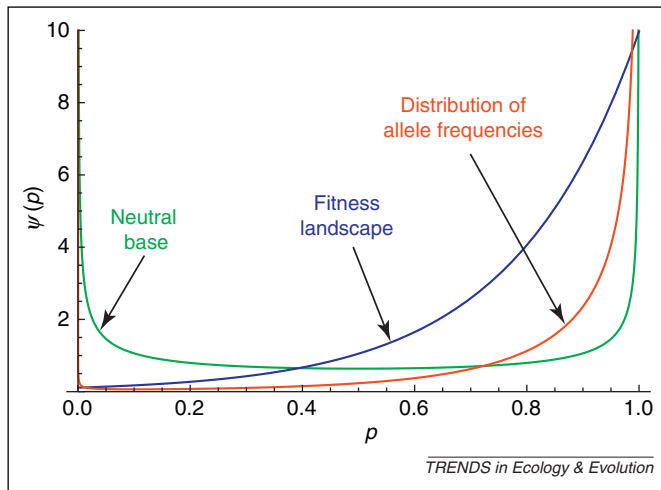


Figure 1. In the solution to the diffusion equation, the effects of fitness (blue) combine with neutral factors (green) to give the distribution of allele frequencies (red). The Metropolis–Hastings algorithm has an analogous structure: the acceptance weights (blue) and the random fluctuations (green) combine to give the distribution that is being estimated (red).

their fitness by accumulation of favourable mutations, or decline under Muller's ratchet [28–30]. Beneficial mutations increase in frequency independently at the wave front, where frequencies are low and subject to drift, but the rest of the wave follows deterministically [31]. The wave thus moves at a velocity proportional to the mutation rate, and which depends logarithmically on the population size because of strong random drift at the leading edge [28]. This approach has been extended to low rates of recombination [32,33]. With sexual reproduction, random drift has less effect, and the population adapts more quickly [30,34]. However, when there is a high rate of substitution and recombination, Hill–Robertson interference limits adaptation rate [35].

Spatial evolution and range expansions

Fisher introduced a simple nonlinear diffusion equation describing the spread of a beneficial mutation through

space [36]. Although motivated by an evolutionary problem, this model raised interest among physicists and mathematicians (establishing a subdiscipline studying the Fisher–KPP model, for Kolmogorov–Piskounov–Petrovski, co-discoverers of the model [37]). Travelling waves explain the decreased genetic diversity that arises from hitchhiking at the leading edge [38,39]. This approach also provides a practical way to measure selection coefficients [40] and, perhaps, a means to distinguish fixation owing to selective sweeps from simple drift [41,42].

Statistical mechanics and the quantitative genetics of finite populations

Although the diffusion equation provides an exact description of evolution, the joint distribution at many loci is hard to grasp. Statistical mechanics simplifies the problem by following just a few variables that summarize all the allele frequencies (or in physics, the states of the particles). These map the fitness landscapes for allele frequency onto a simpler one for quantitative traits [43], which are analogous to macroscopic quantities in statistical physics (Box 2; Figure 2) [44].

Maximization of entropy

This reduction in dimensionality requires a way to account for the degrees of freedom lost in averaging over the underlying genetic states. This can be achieved by applying the principle of entropy maximization: one assumes that the unknown microstates follow a distribution that maximizes their entropy, S , given the values of macroscopic quantities [45]. Entropy can be defined in several ways. The definition appropriate here is analogous to the above, but extends to the case when \underline{p} (the vector of allele frequencies at each locus) is the random variable: $S = - \int \psi \log[\psi/\varphi] d\underline{p}$. This defines the dispersion of the distribution of allele frequencies, ψ , relative to a base distribution, φ : it is maximized when the distribution is selectively neutral ($\psi = \varphi$) and decreases as the distribution becomes more tightly

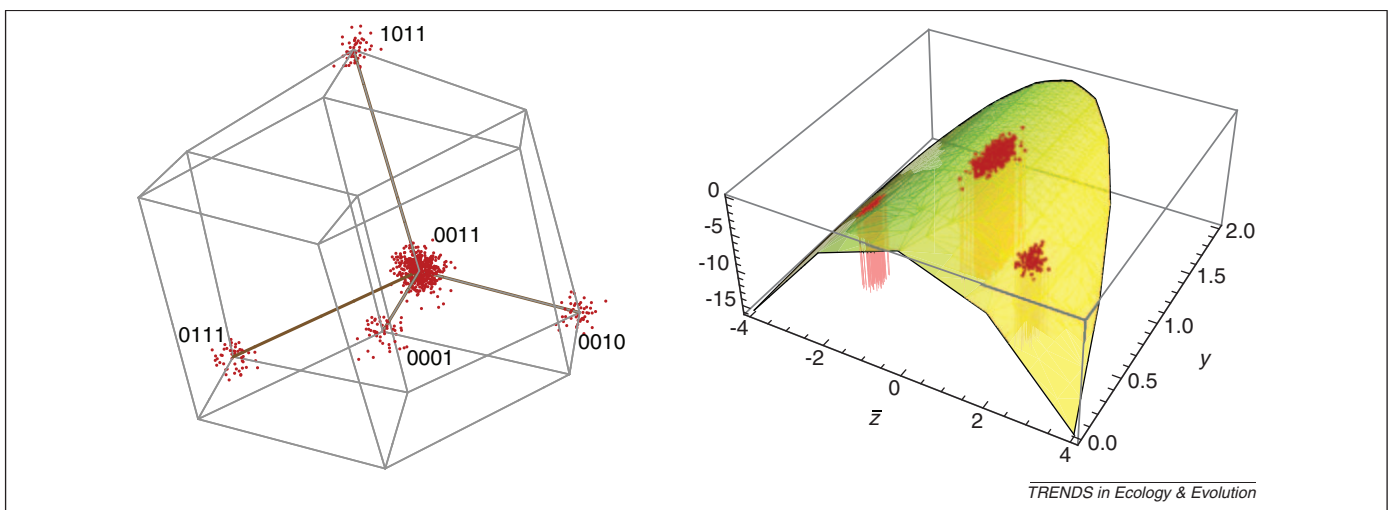


Figure 2. Mapping the genetic fitness landscape to a quantitative-trait fitness landscape. (a) Different combinations of allele frequencies lie in a hyperspace (shown only for a projection of four loci), where the axes represent the frequency of each allele. In this plot, each point represents a population. The dense cloud of points towards the centre is an optimal peak, set at 0011. The other clouds are at suboptimal adaptive peaks one mutation away from the optimum. However, each genotype determines a trait, and the population is mapped to a space of trait means, z , and genetic variance, v . Thus, mean fitness, trait mean and genetic variance, although related by the allele frequencies, generate a fitness landscape in quantitative variables [(b) yellow surface, the height indicating log-mean fitness]. The number of variables (degrees of freedom) is collapsed from a hyperspace of an arbitrary number of allele frequencies at each locus to two quantitative variables: trait mean and genetic variance.

clustered around states that are a priori improbable [46–48]. If S is maximized while constraining the expectations of some macroscopic variables, $\langle A_i \rangle = \int A_i \psi d\mathbf{p}$, we obtain a distribution of allele frequencies $\psi = Z^{-1} \exp[2N \sum_i \alpha_i A_i]$ [46], where Z normalizes the distribution and N is the population size. Remarkably, this distribution corresponds exactly to the stationary solution of the diffusion equation (Box 2), when the A s are chosen according to the particular mode of selection (quantitative traits, genetic variance, etc.) and heterozygosity, and are conjugated with the α s, which are the selection coefficients, mutation rates, and so on [46–48]. This analogy between statistical mechanics and evolution of a finite population has yielded several results, of which we mention a few below.

The dynamics of polygenic evolution can be approximated by a quasi-equilibrium assumption; that is, that the transient distribution of allele frequencies behaves as though the entropy is maximized at all times, given the current values of macroscopic variables. In this way, the change through time of quantitative characters (including their genetic variance) can be computed for populations affected by mutation, selection and drift, for an arbitrary number of loci [46,49]. In physics, macroscopic systems often change more slowly than the microscopic fluctuations, justifying this approximation. In biology, there is not such a stark separation. Nevertheless, the approximation is remarkably accurate even when the environment changes abruptly [46,47,49]; travelling waves might provide an explanation [31].

Adaptive landscapes and detailed balance

Wright's formula for the stationary distribution [20] requires detailed balance [50]. Population geneticists have shown that detailed balance is generally violated when there are more than two alleles at a locus [20], when recombination or migration are comparable with the strength of selection, or under frequency-dependent selection [51]. Without detailed balance, the dynamics cannot be represented by an adaptive landscape, and can be mathematically intractable (although see [52]). Phylogenetic analysis reveals deviations from detailed balance; for example, when genomic GC content changes over time [53]. Therefore, methods are needed for analysing populations that are in a stationary state that violates detailed balance, or that are not at a statistical equilibrium at all.

Path ensembles

An alternative method that holds without detailed balance is the path ensemble [24,54]. Instead of describing the distribution of allele frequencies at any single time, the distribution of paths of allele frequencies is followed between two states at different time points (Figure 3). The probability of any path can be written down in a simple form, and the chance of a transition from one state to another obtained (in principle) by integrating over paths (Box 2). The trajectories are weighted with respect to an optimal one, through three terms: Fisher's information, the variance in fitness and the fitness flux, ϕ (Box 3) [54]. The latter measures the net amount of adaptation given the history of a population. It is defined as $\phi = s \frac{dp}{dt}$, where s is the selective coefficient of the beneficial allele; ϕ is the

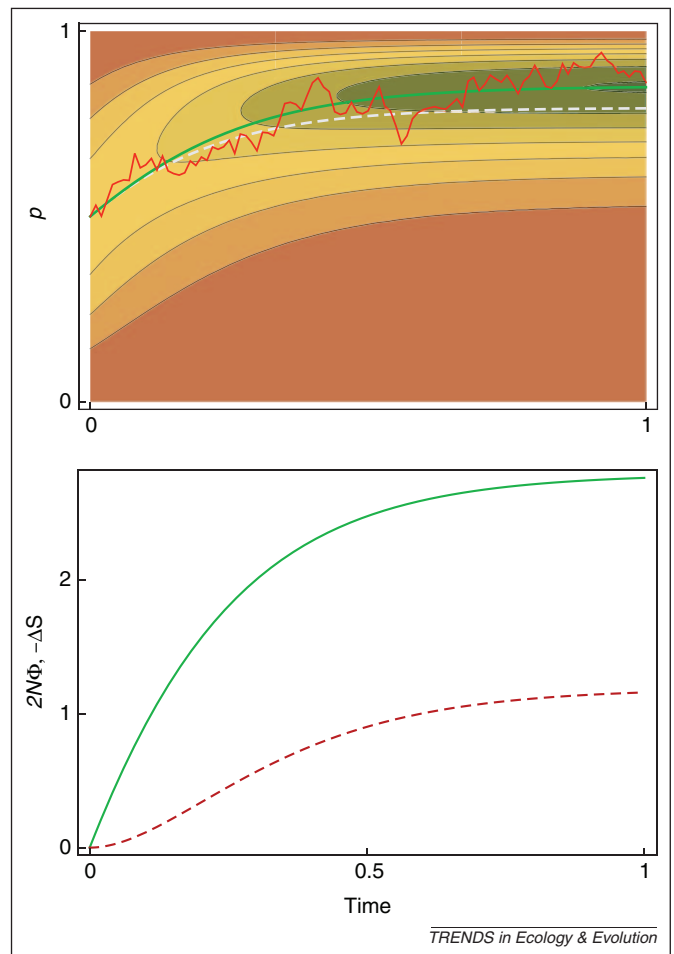


Figure 3. The path ensemble and the distribution of histories. (a) shows the distribution of allele frequencies through time (shown as contour levels). Initially, populations follow the neutral distribution (left axis; $N\mu = 0.7$). Directional selection, $Ns = 2.5$, is then applied, and populations settle to a new distribution (right axis). Any actual realization (red curve) fluctuates stochastically around an optimal one (illustrated with the green curve). The white dashed line is the deterministic solution, shown as a reference. (b) shows the fitness flux (upper curve, green) and the decrease in entropy (lower curve, red). When selection changes abruptly, as here, fitness flux is substantially greater than the decrease in entropy. However, if selection were to change slowly, the two would be equal throughout.

increase in mean fitness that is expected from changes in allele frequency, but without allowing for changes in selection. The fitness flux is distinct from the change in mean fitness, which in general is not well defined when selection changes through time. Fitness flux includes changes in allele frequencies owing to all evolutionary processes and, to the extent that these interfere with selection, can be negative. In considering the history of a population, the path ensemble methods give an understanding of the adaptation and evolution of complex traits that accounts for historical contingencies, an advantage over models that only consider the state of a population at a given time.

Evolutionary biology and Monte-Carlo methods

Monte Carlo methods are now widely used in statistical inference. When many variables are involved, it is not feasible to explore the whole space of possible states (e.g. all possible phylogenetic trees among multiple species). A group working on nuclear weapon development at Los Alamos introduced a simple but widely used algorithm [55]. One simply makes a random change to the microscopic

Box 3. Path ensembles and fitness flux

A path ensemble considers all the possible histories of a population between two fixed states p_0 and p_T at times 0 and T . In this description, each history is the variable being described. The probability of a particular trajectory $p(t)$ is proportional to the factor $\exp\left[-N \int \left(\frac{dp}{dt} - M_{sp}\right)^2 \frac{dt}{p(1-p)}\right]$, where the allele frequencies p are evaluated at each point of the history $p(t)$. Here, M_{sp} is the same factor in the diffusion equation, suggesting the connection between the two methods. The path integral can be understood as a sum if the history is sampled at discrete times, $p = \{p_0, p_1, \dots, p_T\}$. Notice that, because the integral is always positive, if it achieves a minimum for a given history, then that history has the highest probability. To understand the meaning of the integral, we can develop the binomial expression inside the integral into three terms: $F - 2\Phi + v$, and consider the case of selection, $M_{sp} = p(1-p)s$; F is Fisher's information, $\Phi = \frac{M_{sp}}{p(1-p)} \frac{dp}{dt} = s \frac{dp}{dt}$ is the fitness flux and v is the additive genetic variance in fitness. Thus, the histories occur as a compromise between minimizing Fisher's information and genetic variance (both regarded as measures of the speed of adaptation) and maximizing the fitness flux.

The fitness flux is a measure of adaptation of beneficial alleles [54], whereas the cumulative flux ($\Phi = \int \phi dt$) of a population history is the equivalent measure to the fitness of a population (if we think of successive substitutions, it is the total of all the selection coefficients associated with each substitution). The expectation of cumulative fitness flux is necessarily greater than the reduction in entropy between the initial and final equilibrium states (which can be understood as the information gained by the population) [49]:

$2N\langle\Phi\rangle \geq -\Delta S$. That is, it takes a certain amount of selection (measured precisely by the fitness flux) to move the allele frequency distribution away from its neutral state (as measured by the decrease in entropy). This result is generally valid and is not restricted to, say, constant selection. Moreover, if selection changes slowly so that the distribution stays close to the stationary state, then $2N\langle\Phi\rangle = -\Delta S$; such changes are termed 'reversible'. For example, assume that the allele frequencies initially follow a neutral distribution (mutation–drift balance). Suddenly, directional selection is applied so that $W = \exp(sp)$, and loci move toward a new distribution under selection and drift. The fitness flux is then substantially greater than the decrease in entropy (Figure 3, main text). If, by contrast, selection were increased very slowly, eventually to reach the same strength, the net fitness flux would necessarily be much smaller, and equal to the decrease in entropy (lower curve in Figure 3, main text). The fitness flux method is surprisingly general. However, its relationship with quantities that might actually constrain the extent of selection remains obscure. In particular, the additive variance in fitness is proportional to $s^2 p(1-p)$; we see that the additive genetic variance in fitness is just twice the fitness flux, when that includes only the change in allele frequency owing to selection, $\Delta sp = sp(1-p)$. Further understanding could emerge relating the decrease in entropy owing to selection to the additive genetic variance in fitness [90]. Last, it is relevant that fitness flux presents an extension of Fisher's Fundamental Theorem of Natural selection [71]: it considers not only the change owing to selection, but also the effects of drift and, unlike Fisher's theorem, the fitness flux theorem holds also for weak selection ($Ns \sim 1$) [54].

variables, accepting it if it increases some measure, L (e.g. mean fitness). Changes that decrease L to L^* are accepted with probability L^*/L . This ensures that the microscopic variables will follow a distribution proportional to the stationary distribution of the diffusion equation, which would, in turn, be determined solely by the random changes, multiplied by L (Figure 1). This Metropolis algorithm has been developed in a statistical context [56], and applied to generate likelihood surfaces for statistical inference [57–59]. Intriguingly, this algorithm uses a simple form of selection to generate a distribution equal to the product of a neutral base distribution, and the measure L , just as selection and random drift lead to Wright's distribution under the diffusion approximation (Box 2). Both rely on detailed balance, but a path ensemble approach allows extension to more general cases [24].

Obstacles to overcome

Toy models and method-oriented analyses

Over the past decade, physicists have shown strong interest in evolution. For example, during the past 5 years, over 2000 publications on evolution appeared in physics journals (chiefly *Physical Review* journals, *Physica A* and *Proceedings of the National Academy of Sciences USA*). Unfortunately, most of these works pay little attention to the fundamental biology, because the motivation is often the specific methods rather than the biological questions. Consequently, many of these contributions remain unconnected to the rest of the evolutionary theory; for the most part, there is little communication between the disciplines. Two examples follow.

In the Bak–Sneppen model [60], populations evolve by removing the least-fit individual together with two unrelated neighbours, and replacing them with three new individuals with random fitness. A 'critical value' is

reached, but with repeated periods where the fitness distribution spreads, and then reorganizes to the critical value. The Bak–Sneppen model attempted to explain the distributions of extinction episodes [61] and patterns of experimental evolution [62], but had little impact in biology because it lacked any mechanistic basis. Notably, only 13 of 700 citations of the Bak–Sneppen model [60] were by non-physicists. Second, in the Penna bit-string model of ageing [63], the position in the genome of an allele dictates the age at which its detrimental effect is expressed. A threshold for the total number of such deleterious mutations is set arbitrarily and the population evolves under mutation and competition. Senescence arises because selection is less effective in late life, a phenomenon already well understood from Hamilton's general analysis [64]. Here, out of approximately 230 citations to [63], only five did not include physicists. These two approaches, and others alike, are not taken seriously because they rest on 'toy models' that are not connected with biological reality.

Two problems that restrict communication between disciplines

First, the language and nomenclature used by physicists are often not consistent with basic concepts in genetics: they use terms such as energy, spin glass, magnetization, Ising chain, and so on, where they should use mean fitness, polygenes, directional selection or polygenic trait [47,65–67]. Standard population genetics notation is largely ignored, making even the most basic equations appear unfamiliar. To take a central example, the diffusion equation includes deterministic and stochastic 'forces'. In evolution, the stochastic part models genetic drift. However, the term 'drift' is used in physics to refer to the deterministic part. Different nomenclatures make it difficult for physicists to

address important biological questions, and for biologists to understand the questions posed by physicists. This is amplified when new ideas are introduced. For example, in an explanation of the advantages of sex, the idea of mixability was introduced [68]: that is, sex favours alleles that are fit across different genetic backgrounds. A recently proposed measure of mixability [68] is identical to Fisher's analysis of variance, which was devised precisely as a measure of epistasis [69]. Take another example: a statistical mechanics approach was used to find the distributions of contributions made by individual ancestors to future generations. This defined the statistical 'weight' of the contribution of each individual in a lineage [70], which, in biological terms, is just the reproductive value of an individual; again, a concept introduced by Fisher [71].

Second, known results are often rediscovered owing to the lack of a common language. For example, the original result that free fitness increases in evolution was illustrated with several examples from population and quantitative genetics, and was interpreted in terms of selection and drift [48]. Yet, the same principle was twice rediscovered by physicists decades later but with more restricted scope [50]. Another example is the NK model, where the fitness landscape can be 'tuned', altering the degree of epistasis for fitness; this model was used to show that recombination is an evolvable trait [72]. Yet, the theoretical analysis of the evolution of sex and recombination has been a thriving field since the 1970's [73]. No doubt population geneticists have rederived results that are well known in physics (e.g. Wright's calculation of rates of shift between adaptive peaks), but these are not usually published as new physics, and are typically studied for their biological implications. Nevertheless, physicists have also had a serious commitment to subjects meaningful to evolution. Significant works include those discussed in this article, clonal interference in asexuals [29,32,33,38], an application of percolation theory to speciation [74], extending Haldane's principle to a multilocus trait with partial dominance, epistasis and sexual reproduction [65,66], and ecological explanations of replicator dynamics [75,76]. All these are aimed directly at a biological audience, published in appropriate journals. Generally, physicists often have a sharp intuition about their models, which helps greatly in finding solutions.

Statistical physics is based on universal physical laws. By contrast, biological concepts are relative, plastic, or even arbitrary (e.g. mean fitness or traits). Hence, the analogies with statistical mechanical models are limited, depending on the nature of epistasis, physical linkage of the genes, unpredictable fluctuating selection, and so on. Moreover, there are different ways in which precise analogies can be drawn, limiting their scope: some factors act deterministically (e.g. selection) and other stochastically (mutations or drift).

Conclusions

Many of the fundamental processes of both population genetics and statistical physics are described by diffusion. In evolution, it provides a common framework for features such as the change in allele frequencies [20,77], genealogies [78] and spatial dispersal [36]. All these, and others,

can benefit from methods of nonequilibrium statistical mechanics, which is a major, active field in physics.

The concept of a path ensemble is especially useful, shifting the paradigm from tracking frequencies at each point in time, to considering selection over the whole history of the alleles [78]. This can be applied to both deterministic [79] and stochastic evolution [24]. In turn, long-standing questions about the efficiency of natural selection in building complex phenotypes [3,5,6], and evolution under fluctuating selection, can be readdressed.

Of course, we can ask whether the mathematical paraphernalia that we advocate is of any practical use. Although we should not take mathematical models too literally, they are useful both for generating hypotheses about evolution, and for making sense of ecological and genetic data. Most notably, the neutral theory provides the conceptual framework for analyses of sequence data [22], and quantitative genetics predicts the effects of selection on complex traits [80]. Ideas from statistical mechanics might help by providing new ways to describe the evolution of complex traits, and by suggesting constraints on the efficacy of selection. A clearer understanding of concepts such as fitness flux and entropy suggest new ways to think about the evolution of quantitative traits. To understand adaptation, researchers need to contemplate not only the current state of populations, but also their history. This is of course an old idea, but the rationale that we review suggests new ways to understand the process of adaptation in a historical and quantitative way.

Acknowledgements

We would like to thank J.P. Bollback, R. Cipriani, J. Hermisson, J. Polechova and D. Weissman for their comments and observations. This research was funded by the ERC-2009-AdG Grant for project 250152 SELECTIONINFORMATION.

References

- 1 Boltzmann, L. (1896) *Vorlesungen über Gastheorie*, J.A. Barth
- 2 Callen, H.B. (1985) *Thermodynamics and an Introduction to Thermostatistics*, John Wiley & Sons
- 3 Maynard-Smith, J. (2000) The concept of information in biology. *Philos. Sci.* 67, 177–194
- 4 Shannon, C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423
- 5 Haldane, J.B.S. (1957) The cost of natural selection. *Genetics* 55, 511–524
- 6 Kimura, M. (1961) Natural selection as the process of accumulating genetic information in adaptive evolution. *Genet. Res.* 2, 127–140
- 7 Worden, R.P. (1995) A speed limit for evolution. *J. Theor. Biol.* 176, 137–152
- 8 Eigen, M. and Schuster, P. (1977) The hypercycle: a principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541–565
- 9 Kimura, M. and Maruyama, T. (1966) The mutational load with epistatic gene interactions in fitness. *Genetics* 54, 1337–1351
- 10 Frieden, B. *et al.* (2001) Population genetics from an information perspective. *J. Theor. Biol.* 208, 49–64
- 11 Frank, S. (2009) Natural selection maximizes fisher information. *J. Evol. Biol.* 22, 231–244
- 12 Crow, J. and Kimura, M. (1964) The theory of genetic loads. *Proc. 11th Intern. Congr. Genet.* 3, 495–506
- 13 Ewens, W.J. (1979) *Mathematical Population Genetics*, Springer-Verlag
- 14 Mackay, D.J.C. (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press
- 15 Peck, J. and Waxman, D. (2010) Is life impossible? Information, sex and the origin of complex organism. *Evolution* 64, 3300–3309

- 16 Haldane, J.B.S. (1937) The effect of variation in fitness. *Am. Nat.* 72, 337–349
- 17 Watkins, C.J.C.H. (2002) The channel capacity of evolution: ultimate limits on the amount of information maintainable in the genome. *Proc. 3rd Intern. Conf. Bioinf. Genome Reg. Struct.* 2, 58–60
- 18 Fisher, R.A. (1922) On the dominance ratio. *Proc. R. Soc. Edinb.* 42, 321–341
- 19 Kolmogorov, A.N. (1935) Deviations from Hardy's formula in partial isolation. *C.R. Acad. Sci. U.R.S.S.* 8, 129–132
- 20 Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16, 97–159
- 21 Kimura, M. (1955) Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb. Symp. Quant. Biol.* 20, 33–53
- 22 Kimura, M. (1985) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
- 23 Wright, S. (1988) Surfaces of selective value revisited. *Am. Nat.* 131, 115–123
- 24 Barton, N.H. and Rouhani, S. (1987) The frequency of shifts between alternative equilibria. *J. Theor. Biol.* 125, 397–418
- 25 Wright, S. (1941) On the probability of fixation of reciprocal translocations. *Am. Nat.* 75, 513–522
- 26 Lande, R. (1985) The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* 54, 323–332
- 27 Rouhani, S. and Barton, N.H. (1987) Speciation and the shifting balance in a continuous population. *Theor. Popul. Biol.* 31, 465–492
- 28 Wilke, C.O. (2004) The speed of adaptation in large asexual populations. *Genetics* 167, 2045–2053
- 29 Rouzine, I.M. *et al.* (2007) The travelling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Pop. Biol.* 73, 24–46
- 30 Bürger, R. (1999) Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* 153, 1055–1069
- 31 Hallatschek, O. (2011) The noisy edge of travelling waves. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1783–1787
- 32 Rouzine, I.M. and Coffin, J.M. (2010) Multi-site adaptation in the presence of infrequent recombination. *Theor. Popul. Biol.* 77, 467–481
- 33 Neher, R.A. *et al.* (2010) Rate of adaptation in large sexual populations. *Genetics* 181, 467–481
- 34 Peck, J.R. and Waxman, D. (1999) Sex and adaptation in a changing environment. *Genetics* 153, 1041–1053
- 35 Barton, N.H. (2009) Why sex and recombination? *Cold Spring Harb. Symp. Quant. Biol.* 74, 158–170
- 36 Fisher, R. (1937) The wave of advance of advantageous genes. *Ann. Eugen.* 7, 355–369
- 37 Kolmogorov, A.N. *et al.* (1991) A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. In *Selected Works of A.N. Kolmogorov* (Tikhomirov, V.M., ed.), pp. 248–270, Kluwer
- 38 Hallatschek, O. and Nelson, D.R. (2008) Gene surfing in expanding populations. *Theor. Popul. Biol.* 73, 158–170
- 39 Ralph, P. and Coop, G. (2010) Parallel adaptation: one or any waves of advance of an advantageous allele? *Genetics* 186, 647–668
- 40 Bauer, G.J. *et al.* (1989) Travelling waves of *in vitro* evolving RNA. *Proc. Natl. Acad. Sci. U.S.A.* 20, 7937–7941
- 41 Korolev, M. *et al.* (2010) Genetic demixing and evolution in linear stepping stone models. *Rev. Mod. Phys.* 82, 1691–1718
- 42 Hallatschek, O. and Nelson, D.R. (2010) Life at the front of an expanding population. *Evolution* 64, 193–206
- 43 Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30, 314–334
- 44 Rattray, M. and Shapiro, J. (2001) Cumulant dynamics of a population under multiplicative selection, mutation, and drift. *Theor. Popul. Biol.* 60, 17–32
- 45 Prügel-Bennett, A. (1997) Modelling evolving populations. *J. Theor. Biol.* 185, 81–95
- 46 Barton, N.H. and de Vladar, H.P. (2009) Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics* 181, 997–1011
- 47 Barton, N.H. and Coe, J. (2009) On the application of statistical physics to evolutionary biology. *J. Theor. Biol.* 259, 317–324
- 48 Iwasa, Y. (1988) Free fitness that always increases in evolution. *J. Theor. Biol.* 135, 265–281
- 49 de Vladar, H.P. and Barton, N.H. (2011) The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *J.R. Soc. Interface* 8, 720–739
- 50 Sella, G. and Hirsh, A. (2005) The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9541–9546
- 51 Taylor, C. *et al.* (2006) A symmetry of fixation times in evolutionary dynamics. *J. Theor. Biol.* 243, 245–251
- 52 Ao, P. (2008) Emerging of stochastic dynamical equalities and steady state thermodynamics from darwinian dynamics. *Commun. Theor. Phys.* 49, 1073–1090
- 53 Galtier, N. *et al.* (2001) GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159, 907–911
- 54 Mustonen, V. and Lassig, M. (2010) Fitness flux and ubiquity of adaptive evolution. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4248–4253
- 55 Metropolis, N. *et al.* (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092
- 56 Hastings, W. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97
- 57 Szymura, J. and Barton, N.H. (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *Bombina variegata*, near Cracow in southern Poland. *Evolution* 40, 1141–1159
- 58 Geyer, C.J. and Thompson, E.A. (1992) Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J.R. Statist. Soc. B* 54, 657–699
- 59 Beaumont, M.A. (2010) Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* 41, 379–406
- 60 Bak, P. and Sneppen, K. (1993) Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.* 71, 4083–4086
- 61 Sneppen, K. *et al.* (1995) Evolution as a self-organized critical phenomenon. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5209–5213
- 62 Elena, S.F. and Sanjuán, R. (2005) RNA viruses as complex adaptive systems. *Biosystems* 81, 31–41
- 63 Penna, T.J.P. (1995) A bit-string model for biological aging. *J. Stat. Phys.* 78, 629–1633
- 64 Hamilton, W. (1966) Moulding of senescence by natural selection. *J. Theor. Biol.* 12, 12–45
- 65 Baake, E. *et al.* (1997) Ising quantum chain is equivalent to a model of biological evolution. *Phys. Rev. Lett.* 78, 559–562
- 66 Baake, E. and Wagner, H. (2001) Mutation–selection models solved exactly with methods of statistical mechanics. *Genet. Res.* 78, 93–117
- 67 Hermisson, J. *et al.* (2002) Mutation–selection balance: ancestry, load and maximum principle. *Theor. Popul. Biol.* 62, 9–46
- 68 Livnat, A. *et al.* (2010) Sex, mixability, and modularity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1452–1457
- 69 Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433
- 70 Derrida, B. *et al.* (2000) Distribution of repetitions of ancestors in genealogical trees. *Physica A* 281, 1–16
- 71 Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*, (1st edn), Clarendon Press
- 72 Kauffman, S.A. (1993) *The Origins of Order*, Oxford University Press
- 73 Maynard-Smith, J. (1978) *The Evolution of Sex*, Cambridge University Press
- 74 Gavrillets, S. (1997) Evolution and speciation on holey adaptive landscapes. *Trends Ecol. Evol.* 12, 307–312
- 75 Demetrius, L. (1983) Statistical mechanics and population biology. *J. Stat. Phys.* 30, 709–753
- 76 Demetrius, L. and Ziehe, M. (2007) Darwinian fitness. *Theor. Popul. Biol.* 72, 323–345
- 77 Wright, S. (1937) The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. U.S.A.* 23, 307–320
- 78 Barton, N.H. and Etheridge, A. (2004) The effect of selection on genealogies. *Genetics* 166, 1115–1131
- 79 Leibler, S. and Kussell, E. (2010) Individual histories and selection in heterogeneous populations. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13183–13188
- 80 Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer Associates
- 81 Price, G.R. (1972) Fisher's fundamental theorem made clear. *Ann. Hum. Genet.* 36, 129–140

- 82 Ewens, W.J. (1989) An interpretation and proof of the fundamental theorem of natural selection. *Theor. Popul. Biol.* 36, 167–180
- 83 Timofeëff-Ressovsky, N. *et al.* (1935) Über die Natur der Genmutation und der Genstruktur. *Nachr. Ges. Wiss. Göttingen* 1, 189–245
- 84 Luria, S. and Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28, 491–511
- 85 Schrödinger, E. (1944) *What is Life?* Cambridge University Press
- 86 Judson, H.F. (1998) *The Eighth Day of Creation: Makers of a Revolution in Biology*, Plainview
- 87 Davis, M. and Etheridge, A. (2006) *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance*, Princeton University Press
- 88 Kolmogorov, A.N. (1931) On the analytical methods in probability calculations. *Math. Ann.* 104, 415–458
- 89 Kimura, M. (1964) Diffusion models in population genetics. *J. Appl. Probab.* 1, 177–232
- 90 Barton, N.H. and Partridge, L. (2000) Limits to natural selection. *Bioessays* 22, 1075–1084