Biological evolution through mutation, selection, and drift: An introductory review

Ellen Baake[1] and Wilfried Gabriel[2]

Zoologisches Institut, Universität München, Luisenstr. 14, D-80333 München, Germany.
[1]*email: baake@zi.biologie.uni-muenchen.de*
[2]*email: wilfried.gabriel@lrz.uni-muenchen.de*

Contents

1 Introduction

Why a review on biological evolution in the Annual Review of Computational Physics? The questions raised are not, in the first place, physical ones, and the results reviewed are only partly computational. However, the past few years have seen a boost of activities in physics directed towards biology, and expectations run high that cooperation between biology and physics will constitute a flourishing branch of science in the next century [1].

Physicists seem particularly attracted by the field of biological evolution. The topic lends itself readily to application of concepts, analytical tools, and numerical methods from statistical physics, with biological understanding profiting from this new viewing angle. However, *population genetics*, which describes biological evolution in mathematical terms, is a venerable discipline, which was founded by Fisher, Haldane, and Wright in the twenties (R.A. Fisher may be better known to physicists as the father of modern statistics.). The field has developed to a high degree of sophistication and is laid down in a far-spread-out literature, much of which is hard to access even for biologists, and much more so for physicists. Hence, there is a tendency to run into avoidable pitfalls and to reinvent the wheel. Even worse, well-established concepts are sometimes misunderstood and redefined in misleading ways, and redundant terminology and notation is being created in the course of action. This considerably adds to entropy, and impedes communication. The purpose of the present review is therefore twofold. We shall first review some population genetic foundations in a nutshell. Since the field is rather mature, many pointers will be set to other reviews or even books. We will then bring some structure into the emerging field by bundling up recent work on mutation-selection models. But first of all, we must define our subject more precisely.

'Evolution proceeds via mutation and selection.' In this shortened version, the understanding of biological evolution has entered everyday knowledge. Of course, it's too short to be correct: actually, evolution consists of a fair number of elementary processes, i.e. mutation, selection, recombination, migration, and drift (i.e. fluctuations due to finite population size), and it is not even decided whether mutation and selection are the most important ones. They are, however, weighty factors, and may serve as a case study of evolution as such. For the conceptual issues related to the notion of evolution, we refer the reader to Endler [40].

We shall restrict ourselves to models of mutation and selection, without and with genetic drift. That is, we must – regrettingly – exclude the fascinating topic of recombination, and, with it, the evolution and maintenance of sexual reproduction. Fortunately, there is a recent review available on this topic [43]. But due to finite time, space, and knowledge of the authors, the field must be narrowed down even further. We shall concentrate on models which describe *genetic variation within single populations*. This excludes macroevolution, speciation, and phylogeny (which are concerned with variation *between* populations), as well as adaptive walks (which describe species as *genetically homogeneous* entities). For the same reasons, models of coevolution will not be considered here. For a simple model of interacting populations, we refer to Bagnoli and Bezzi (this volume).

In order to provide the uninitiated with some flavour of the field, we will set out with a few relevant questions before embarking on the basic models of population genetics. The latter describe the evolution of the composition of a population under the joint action of mutation, selection, and drift, and will provide the foundations on which to build. Since deleterious mutations and mutational degradation have been major concerns in the context of asexual species, we shall then dwell on the major phenomena that have been considered in this context, namely error thresholds, Muller's ratchet, and mutational meltdowns. We shall finally turn to recent developments which, on the one hand, show ways out of mutational degradation, and, on the other hand, connect to dynamical aspects of evolution.

We shall be mainly concerned with models aiming directly at the genetic (as opposed to the phenotypic) level, thereby setting aside much of quantitative genetics. Mutation-selection models with an emphasis on quantitative genetics have been recently covered by a review [19], and a textbook [118]. We strongly recommend the simultaneous use of these and other reviews (which will be mentioned while we proceed), as well as the standard textbooks of population genetic theory. Among them are the four volumes of Wright's 'classic' [175, 176, 177, 178], Crow and Kimura (1970) [27], Ewens (1979) [41], and Nagylaki [133]. Molecular evolution in particular is covered by Ratner et al. (1995) [146]. For the less mathematically minded, Hartl and Clark (1997) [71] (student level book with an excellent bibliography), Li (1997) [113] (comprehensive overview of molecular evolution), and Maynard Smith (1989) [122] (for overwhelming biological intuition) are warmly recommended.

2 Some questions

Much attention has been devoted to questions related to the equilibrium situation known as *mutation-selection balance*. What is the equilibrium composition of the population under the simultaneous action of mutation and selection? How large is the *mutation load*, i.e. the unavoidable loss of fitness due to the production of less-well-adapted individuals by mutation (the 'cost of variability', as it were)? How large is the genetic variability? Is mutation-selection balance sufficient to explain the observed variability? This comparison of data and theory has, in the past, led to crucial insights both in classical and molecular genetics, for review, see [12, 60].

Apart from such equilibrium considerations, dynamical aspects have become more and more important. Can present-day genetic variation be used to draw conclusions about the (past) operation of selection? In particular, inference of selection at the molecular level is currently a 'hot topic' (cf. [163, 84, 48] for a few original landmark papers, [6, 83] for reviews, and [113, Ch. 9] for a comprehensive overview.)

Often, the joint action of mutation, drift, and selection leads to a steady state with a stationary phenotype under an ongoing turnover of genotypes. What are the rules that

govern this turnover? The latter question is very important for between-population comparisons and also touches on *molecular phylogeny*, the art of reconstructing evolutionary trees from sequence data; for review, see [160] and [113, Ch. 5 and 6].

3 Mutation, selection, and drift

3.1 Mutation and selection: The basics

The basic equations: Loosely speaking, a *gene* is a portion of the genome which codes for something, for example an enzyme. When special reference is made to the location (the *site*) in the genome, one speaks of the gene *locus*. A gene may occur in several versions called *alleles*.

Let us, for the moment, consider a population of haploid organisms which carry but one set of genes or chromosomes per cell, like viruses, bacteria, or blue-green algae, which reproduce asexually. We shall focus on one single gene with K alleles, $A_1, \dots, A_K$, which will sometimes be called 'types' in what follows. It is further assumed that fitness is solely determined by the allele at the locus in question, i.e. variability at all other loci will be ignored. Individuals may then be characterized by, and identified with, alleles.

Let us first assume that generations are discrete and nonoverlapping. Individuals are counted at the beginning of every new generation ('before selection') and undergo the simple process

$$A_i \xrightarrow{v_{ji} w_i} A_j \, . \tag{1}$$

Every A_i individual produces, on average, $w_i \geq 0$ offspring for the next generation, and then dies; included in w_i are the probability of survival to the reproductive age (the *viability*), and the number of offspring (the *fecundity*). At every reproduction event, mutation may occur, i.e. the offspring of A_i is of type A_j with probability v_{ji}. The mutation matrix $\mathcal{V} = (v_{ij})_{1 \leq i,j \leq K}$ is a Markov matrix, i.e. $v_{ij} \geq 0$ and $\sum_{i=1}^K v_{ij} = 1$. The quantities w_i are known as the *Wrightian fitness values*, cf. [27]. If the average number of offspring is the same for all genotypes and only viability differences are considered, one speaks of *viability fitness*. In both cases, $\mathcal{W}$, the diagonal matrix $\mathrm{diag}(w_1, \dots, w_K)$, may be understood as the reproduction matrix (one should not use the term 'fitness matrix' here, since this is reserved for the fitness of diploid genotypes, as described below). We shall, however, use the notion of *fitness landscape* for the mapping from allelic space to fitness, due to its intuitive appeal as a (high-dimensional) mountain range with a population moving within it like a cloud, 'trying' to access the highest peaks. The picture goes back to Wright (1932) [174], who coined the metaphor of an *adaptive landscape*. However, two different meanings became associated with this expression (cf. [56]), wherefore we prefer the term 'fitness landscape'. Concrete fitness landscapes, as well as mutation models, will be specified lateron.

Let us now consider a *population* of individuals which is so large that the frequencies of the various types in the population may be treated as continuous quantities, and random fluctuations may be neglected. As long as there is no restriction on population size, the change of the composition of the population across generations is then described by the linear difference equation

$$x_i' = \sum_{j=1}^{K} v_{ij} w_j x_j \,, \quad \text{or} \quad \boldsymbol{x}' = \mathcal{V}\mathcal{W}\boldsymbol{x} \,. \tag{2}$$

Here, x_i and x_i' denote the *absolute* frequencies of A_i individuals in successive generations, and $\boldsymbol{x} := (x_1, \ldots, x_K)^T$, where T denotes transpose. The corresponding *relative* frequencies ($p_i := x_i / \|\boldsymbol{x}\|_1$ where $\|\boldsymbol{x}\|_1 = \sum_i x_i$ is the total population size) are, however, more interesting, and more readily observable from population samples. Application of (2) yields the *nonlinear* discrete dynamical system

$$p_i' = \frac{\sum_j v_{ij} w_j p_j}{\sum_j w_j p_j} \,, \tag{3}$$

where $\sum_j w_j p_j =: \bar{w}$ is the *mean (Wrightian) fitness* of the population. Note that $\sum_i p_i = 1$, and, therefore, a probabilistic interpretation is adequate, although the dynamics itself is deterministic.

Eq. (3) is the haploid version of the mutation-selection equation of population genetics as originally formulated by Haldane (1928) [70], reviewed in detail by Crow and Kimura (1970) [27], and recently by Bürger (1998) [19]. We have derived it here for the most basic situation: a haploid population, discrete time, infinite population size (i.e. no genetic drift), and unconstrained population growth. In what follows, these assumptions will be relaxed one by one.

Population dynamics: The mutation-selection equation (3) may seem of limited relevance due to its derivation from unconstrained population growth, which is clearly unrealistic. However, its range of validity is much larger. Consider a scenario where, in addition to genotype-specific reproduction, some kind of population regulation is in effect which eliminates a fraction g of individuals *regardless of their genotypes*, i.e.

$$\bigcirc \xleftarrow{\;g\;} A_i \xrightarrow{v_{ji} w_i (1-g)} A_j \,. \tag{4}$$

Here, eliminated individuals are symbolized by a hole ('$\bigcirc$'). The fraction g may vary from generation to generation; in particular, it may depend on the current population size in a nonlinear fashion. For example, g may constrain population size to a maximal carrying capacity, as in the case of the well-known Verhulst or logistic equations (see, e.g., [132, Ch. 2]). Although the dynamics of absolute frequencies may differ drastically from that predicted by Eq. (2), the dynamics of *relative frequencies* is easily shown to be again described by Eq. (3) for *arbitrary* g, cf. [27]. This goes together with the fact that, in

contrast to Eq. (2), Eq. (3) is obviously invariant under the transformation $w_i \to w_i \cdot c$ for any (positive) constant c, if it is applied to all fitness values simultaneously. That is, *ratios* of fitness values determine the dynamics, rather than absolute values.

As long as populations of moderate size are considered, focusing on relative frequencies seems appropriate. However, this is beside the point when populations get close to extinction; then absolute frequencies *must* be considered. We will come back to this point later.

Basic properties of mutation-selection models: The explicit solution of Eq. (2) is, of course,

$$\boldsymbol{x}^{(n)} = (\mathcal{V}\mathcal{W})^n \boldsymbol{x}^{(0)} \,, \tag{5}$$

where we have used n to indicate generation numbers. From this, the solution of Eq. (3) is obtained by normalization. Without mutation, i.e. $v_{ij} = \delta_{ij}$, Eq. (3) reduces to a haploid version of *Fisher's selection equation*. The explicit solution of this haploid version is trivial, since $\mathcal{W}$ is diagonal. It is well-known that mean fitness acts as a Lyapunov function for the dynamics, i.e. it increases along all trajectories; see, e.g., [41, 79]. This is intuitively obvious since fit individuals flourish at the expense of less fit ones – this is what is called *selection*. As a consequence, only the fittest type(s) (of those initially present) will survive in the long run.

If mutation is present, too, the analytic solvability of the dynamics depends on whether $\mathcal{V}\mathcal{W}$ can be diagonalized explicitly, which is rarely the case for large K. The time evolution may still be determined numerically by van Mises iteration (see, e.g., [181, p. 178]) of (5). If $\mathcal{V}$ is primitive, which is usually the case in biologically relevant situations, existence of and global convergence towards a stationary distribution is guaranteed by the Perron-Frobenius theorem, cf. [87, Appendix]. This stationary distribution is given by the Perron-Frobenius eigenvector of $\mathcal{V}\mathcal{W}$ if supplied with positive sign and correct normalization; cf. [166, 129]. Global convergence also guarantees the stability of the van Mises iteration or related numerical procedures.

3.2 A few extensions

Diploid models: So far, we have only mentioned haploid organisms, i.e. those with only one set of genes per cell. In most 'higher' organisms, however, two copies are present, at least in certain stages of their life cycles; i.e. they are *diploid*. If the phase of the life cycle on which selection acts is haploid (as in algae and mosses), Eq. (3) holds without modification (note that we always exclude recombination). With most species, however, fitness is a function of the diploid genotype. Let w_{ij} be the fitness of an A_iA_j individual, where A_i (A_j) is the allele inherited from the father (mother). Since for most genes (the *autosomal*, as opposed to *sex-linked*, genes), A_iA_j is indistinguishable from A_jA_i, one also has $w_{ij} = w_{ji}$. Eq. (3) must be modified according to the mode of reproduction.

If the diploid genome is passed on to the offspring without reshuffling (as in vegetative reproduction in plants), every genotype A_iA_j may be considered as an entity, and nothing but a simple relabelling is required. If, on the other hand, reshuffling of alleles takes place via random mating and sexual reproduction (note that this does not *per se* imply recombination), alleles are combined independently, i.e. the frequency of A_iA_j is p_ip_j in the next generation; this is the famous *Hardy-Weinberg equilibrium*, cf. [27, 79]. The dynamical equation for the allele frequencies (3) must then be modified by replacing the w_i by the corresponding *marginal fitnesses* $\bar{w}_i := \sum_j w_{ij}p_j$. In contrast to (3), the resulting equations are inherently nonlinear. As a consequence, there may be multiple steady states, or limit cycles (e.g. [2, 78, 10]). Things become simple again, however, if there is no *dominance*. Absence of dominance means that $w_{ij} = \sqrt{w_{ii}w_{jj}}$ for all i and j, i.e. the fitness of every heterozygote is the geometric mean of those of the corresponding homozygotes[a]. Then, the diploid equation reduces to the original haploid one (2) with $w_i := \sqrt{w_{ii}}$ for all i; cf. [79, p.251] and [173]. In order to avoid further complications, we shall, in what follows, adhere to this convenient special case, although we are well aware of the fact that dominance is abundant in diploid organisms, and may change the evolutionary dynamics substantially.

Continuous time: The literature on mutation-selection models is partly in discrete, partly in continuous time. We shall now briefly clarify the relationship between the various models. There are actually *two* continuous-time versions, which correspond to the processes

$$\bigcirc \xleftarrow{d_i} A_i \xrightarrow{v_{ji}b_i} A_i + A_j \tag{6}$$

and

$$\begin{tikzcd} \bigcirc & A_i \arrow[l, "d_i"'] \arrow[r, "b_i"] \arrow[d, "m_{ji}"] & 2A_i\,. \\ & A_j & \end{tikzcd} \tag{7}$$

Here, the b_i and d_i are (instantaneous) birth and death rates, respectively. We are considering approximations of (6) and (7) which are continuous in time with at most one 'event' occurring at any epoch; for example, multiple births may be mimicked by a larger birth rate.

In scenario (6), mutation occurs exclusively on the occasion of reproduction events, with probabilities v_{ji}, as in discrete time. In contrast, mutation is an independent process in (7) and occurs, at rates m_{ji}, at any instant of the life cycle. The difference in notation is to remind us of the fact that the m_{ji} are transition *rates* (and thus have the dimension 1/time), with $m_{ii} = -\sum_{j\neq i} m_{ji}$, whereas the v_{ji} are transition *probabilities* (which are dimensionless), and hence $v_{ii} = 1 - \sum_{j\neq i} v_{ji}$.

[a] We have used the term *dominance* in the sense of *dominance in fitness* here; this should not be confused with *dominance at the phenotypic level*.

As to the reproduction and death rates, $r_i := b_i - d_i$ are the *Malthusian fitness* values. In scenario (6), $d_i \equiv d$ is assumed to keep things simple, whereas this is not required for (7). With this in mind, the differential equations corresponding to (6) and (7) read

$$\dot{p}_i = \sum_j v_{ij}r_jp_j - \bar{r}p_i \tag{8}$$

and

$$\dot{p}_i = (r_i - \bar{r})p_i + \sum_j m_{ij}p_j\,, \tag{9}$$

where $\bar{r} := \sum_j r_jp_j$ is the *mean (Malthusian) fitness* of the population. The first version is sometimes called the *coupled* mutation-selection equation and was studied by Akin [2], Hadeler [68], and others. The second version is the *decoupled* or *parallel* version and seems to appear first in [27]. Biologically, they differ in the assumptions on the mutation mechanism. The coupled equation describes mutation as replication errors, whereas the parallel one understands them as effects of radiation, free radicals or thermal fluctuations. Which is the more relevant contribution in nature is an issue under debate, but undecided. It comes down to the question of whether the number of mutation events is closer to constant in time or constant per generation (see the discussion in, e.g., [63, 60] and [137]). These differences are important for the comparison of species with different generation times, like mice and men. For our within-population framework, however, the distinction is not an important one. Typically, the two continuous as well as the discrete time version give very similar results. For continuous time, this is so because the parallel version emerges from the coupled one in the limit of weak selection and mutation, as was shown by Hofbauer (1985) [78].

The comparison of discrete with continuous time requires a closer look. Comparing the mutation and reproduction operators separately gives

$$\mathcal{V} = \exp(\mathcal{M}\tau) \quad \text{and} \quad \mathcal{W} = \exp(\mathcal{R}\tau), \tag{10}$$

where $\mathcal{M}$ is the mutation matrix with entries m_{ij}, $\mathcal{R} := \mathrm{diag}(r_1, \dots, r_K)$, and τ denotes the duration of one generation. In the limit $n \to \infty$ and $\tau \to 0$ with $n\tau = t = \text{const}$, it follows from the Trotter formula [147] that

$$(\mathcal{V}\mathcal{W})^n = \exp\left(\left(\mathcal{M}\frac{t}{n}\right)\exp\left(\mathcal{R}\frac{t}{n}\right)\right)^n \quad \overset{n\to\infty}{\longrightarrow} \quad \exp((\mathcal{M}+\mathcal{R})t)\,; \tag{11}$$

cf. [169]. Hence, the discrete time equation converges to the parallel one in the *short generation limit*.

At the same time, (10) clarifies the relationship between the Malthusian and Wrightian fitness concepts. Obviously, the invariance of the Wrightian fitness under $w_i \to w_i \cdot c$ translates into invariance under $r_i \to r_i + c$ for Malthusian fitness, i.e. fitness *differences* rather than ratios are relevant here (this also entered the derivations of Eqs. (8) and (9)).

Likewise, other relations change from multiplicative to additive, e.g. the condition for absence of dominance now reads $r_{ij} = (1/2)(r_{ii} + r_{jj})$ for all i and j, where r_{ij} is the Malthusian fitness of genotype A_iA_j.

With this in mind, we may – and shall – move freely between the versions, using the most convenient one for each question considered (or the one authors have happened to find themselves most familiar with).

Finite Populations: Of course, all real populations are *finite*, and this must be taken into account in many situations. Let us therefore consider a haploid population of N individuals, its size remaining constant over generations. The number of offspring per parent will be a fluctuating quantity, however, even in the absence of selection; that is, the (absolute) frequencies $N_1, N_2, \dots, N_K$ of types $A_1, A_2, \dots, A_K$ are *random variables*, subject to the constraint $\sum_i N_i = N$. Actually, the simplest (and most popular) picture is that of parents being sampled with replacement, each with probability $w_i/(\sum_j w_j N_j)$, to give birth to a member of the next generation, which may or may not be mutated. For $w_i \equiv c$ (so-called *neutral evolution*), the number of offspring per individual then follows a binomial distribution with success probability $1/N$, cf. [71, 31]. The resulting fluctuations of allele frequencies are known as *genetic drift*. In general (i.e. with mutation and selection), the transition from one generation to the next is a Markov process defined by multinomial sampling, i.e. the transition probabilities read

$$P(\boldsymbol{n'}|\boldsymbol{n}) = \frac{N!}{n'_1! \dots n'_K!} \psi_1(\boldsymbol{n})^{n_1} \dots \psi_n(\boldsymbol{n})^{n_K} , \tag{12}$$

where $\boldsymbol{n} = (n_1, \dots, n_K)^T$ and $\boldsymbol{n'}$ are the type counts in successive generations, i.e. realizations of the random variable $\boldsymbol{N} := (N_1, \dots, N_K)^T$, and

$$\psi_i(\boldsymbol{n}) := \frac{1}{N} \sum_j v_{ij} \frac{w_j}{\bar{w}} n_j \quad \text{with } \bar{w} := \frac{1}{N} \sum_k w_k n_k \, . \tag{13}$$

This innocent-looking process is known as *Wright-Fisher sampling* and describes the interplay of mutation, selection, and drift. The sampling scheme (12) is quite unwieldy, however, and is better studied through diffusion approximations (at least as long as the number of dimensions is small), see [41]. An alternative is to start from a continuous-time analogue of (12), i.e. a master equation, known as the Moran model [128, 41].

In contrast to the situation with infinite population models, we have, for the sampling process, assumed that some kind of population regulation is in effect to keep population size constant. This goes together with the fact that, like the infinite-population version (3), the Wright-Fisher model is invariant under multiplication of the fitness values with a constant. As long as the reproductive capacity of a population (i.e. $N\bar{w}$) safely exceeds the carrying capacity C of the biotope, this is reasonable. If, however, for some reason or other, $N\bar{w} < C$, absolute (as opposed to relative) fitness values should be taken into account, and population dynamics must be modelled explicitly. After all, in this case the

population may go to extinction, and then this is the one important thing to consider (the gene frequencies in the dying population are irrelevant). We note this here to be kept in mind but defer treatment to later chapters.

Basic properties of the stochastic process: Quite generally, mutation increases genetic variation, whereas selection and drift tend to reduce it. Understanding their simultaneous action quantitatively, however, is tremendously difficult. Below we list a few limiting cases and corner stones; for review, see [176].

If mutation is absent (i.e. $v_{ij} = \delta_{ij}$), the genetically homogeneous states ($N_i = N$ for some i, $N_j = 0$ for $j \neq i$) are absorbing, and every population will finally end up in one of them, an event known as *fixation*. Selection will show up through a higher fixation probability for advantageous alleles. This is more pronounced if population size is large; in small populations, random effects predominate. To be more precise: when a new advantageous (or deleterious) mutant is introduced into an otherwise homogeneous population, its fixation probability is roughly s, provided N is not too small and $Ns > 1$. Here, s is the selective advantage (i.e. the difference in (Malthusian) fitness between mutant and wild type); otherwise, the mutant allele behaves very similar to a selectively neutral one[b]. When more than two different fitness values come into play, the 'trafficking of alleles poses serious difficulties to analysis, and one has to rely on simulations, cf. [61]. We will meet concrete examples later.

If the mutation matrix $\mathcal{V}$ is primitive, the process is strictly ergodic[c], and there is a stationary distribution, which is accessible for a few simple selection schemes [176]. Stationary distributions may be understood as time averages, which makes sense if K is small and N is large, so that all states are visited in evolutionary (or Monte Carlo) time. When the number of possible alleles becomes too large, as is usual in the molecular context, the population will move through the allelic space as a (small) cluster which will 'never' sample the space. Then, the infinite alleles limit ($K \to \infty$) is appropriate [99, 41], and other notions of equilibrium must be sought which refer to the genetic variability within the cluster independently of its position. A lot is known here if selection is absent, i.e. if mutation and drift are the only evolutionary forces. A large edifice of theory, Kimura's *neutral theory of evolution*, has been developed on behalf of this case. It has reached a high degree of mathematical sophistication; for review see [97] or the collection of Kimura's papers annotated by Takahata, [98]. One important measure of within-population variability is the frequency distribution of alleles in a sample without reference to their type; under neutrality, it is given by the celebrated Ewens sampling formula [42, 41]. Parallels between frequency distributions of population genetics and frequency distributions of statistical physics have been reviewed by Higgs [74].

Finite populations, backward in time: A major breakthrough which greatly

[b]Let us mention that, for diploid organisms without dominance, the situation is the same, but N has to be replaced by $2N$. If dominance is present, however, the results may change considerably.

[c]Note that ergodicists call *strictly ergodic* [121] what Markovianists call *ergodic* [87].

advanced both theoretical understanding as well as statistical inference from sequence samples was achieved by Kingman (1982) [102, 103], who considered the evolution of finite populations backward in time. The sampling process (12), which may be considered as a bifurcation process forward in time, is then replaced by a *coalescent process* backward in time; cf. Figs. 1 and 2. The power of the backward process lies in the fact that only lineages surviving to the present need be considered. For the purpose of inference, these are further reduced to those lineages leading up to the sampled individuals. In any case, quantities of prime interest are the statistical properties of the genealogies, i.e. the distribution of coalescence times, and, in particular, of the time back to the most recent common ancestor of a sample of individuals. For neutral evolution, coalescence theory is now well understood and has been reviewed by Hudson [82], and Donnelly and Tavaré [36]. The (relative) accessibility is due to the fact that mutations have no effect on the genealogies because, by definition, neutral mutations do not affect the number of offspring of individuals bearing these mutations. As a consequence, individuals are independent and equivalent with respect to the coalescence process, and the mutation process may be considered separately from the genealogical process.

The basic quantity of the genealogical process is the probability of a *coalescence event* of two or more out of m individuals, $m \ll N$, in a given generation. Noting that all m individuals have distinct ancestors (in the preceding generation) with probability $\prod_{i=1}^{m-1}(1-i/N)$, the coalescence probability is

$$P(m) = 1 - \prod_{i=1}^{m-1}\big(1 - \frac{i}{N}\big) = \frac{1}{N}\binom{m}{2} + O\big(\frac{1}{N^2}\big)\,. \tag{14}$$

The approximation ignores the simultaneous coalescence of more than two lineages, which is reasonable since $m \ll N$. Any two of the present lineages are equally likely to form the coalescing pair. Now, the probability that m individuals have m distinct ancestors in each of the preceding $n-1$ generations, and that a coalescence event takes place in generation n backward in time, is geometrically distributed,

$$Pr(T_m = n) = \big(1 - P(m)\big)^{n-1} P(m)\,, \tag{15}$$

with mean value $P(m) \simeq \frac{1}{N}\binom{m}{2} =: \lambda$. Note that in the limit of continuous time, the geometric distribution approaches the exponential distribution with density function $\lambda \cdot \exp(\lambda t)$. With this in mind, it may be shown that the time back to the most recent common ancestor of a sample of size M, $T_{MRCA} := \sum_{m=1}^{M} T_m$, is exponentially distributed with mean $2N(1-1/M)$.

A neutral mutation process may be superimposed on the coalescence process. Statistical properties of genealogies may then be translated into statistics of genetic similarity. This was exploited by Derrida and Peliti (1991) [31], who rediscovered the (neutral) genealogical process and found it to be equivalent to the annealed random map model from statistical physics [30]. This equivalence allows the convenient calculation of various measures of genetic structure.

As soon as selection comes into play, however, the independence of individuals is lost. Dependence comes in through fluctuations of the mean fitness of the population, which involves *all* individuals. This was long thought an insurmountable obstacle to the treatment of the coalescent process with selection, but a breakthrough has been achieved recently by Neuhauser and Krone (1997) [107, 134] in the framework of interacting particle systems.

Apart from its theoretical importance, the coalescent process is now an indispensable tool for the inference of evolutionary history from sequence samples. The most powerful methods for this purpose involve maximizing the likelihood of the observed sample configuration over the set of parameters of the model class considered. To be more precise, the likelihood of the observed data for a given parameter set Θ (which may specify the mutation process as well as the reproduction process) may be written as

$$L(\Theta) = \sum_{G} Pr(D|G)Pr(G|\Theta), \tag{16}$$

where $Pr(G|\Theta)$ is the probability of the genealogy G given the parameter Θ, and $Pr(D|G)$ is the probability of the data D given the genealogy, see, e.g., [108]. Computation of the overall likelihood along the lines of the coalescent is far less costly than the corresponding simulations forward in time. However, it still demands a summation over a huge number of genealogies. Most of them are so implausible that they contribute almost nothing to the likelihood. Therefore, some kind of importance sampling is indispensable. Different approaches are in use, see [65], [108], and [45] for a review. Still, this kind of inference problem remains demanding from the computational point of view, even for neutral evolution. If selection is considered, too, it becomes a real challenge.

4 Specification of ingredients: Genotypes and phenotypes, mutation models, and fitness landscapes

So far, we have assigned fitnesses to genotypes, thus bypassing the phenotype. Ideally, however, the sequence *genotype* $\rightarrow$ *phenotype* $\rightarrow$ *fitness* should be considered, and we shall do so where possible. It will be apparent, though, that often either the genotype or the phenotype will suffer some neglect – this is inevitable in view of the notorious inaccessibility of the genotype-phenotype mapping. To be more precise, a certain tradeoff will be observed. If modelling is aimed at the genotypic level, one may easily formulate mutation models which are plausible in molecular terms; however, the mapping from genotype to fitness is necessarily artificial. If, on the other hand, the phenotype is in the centre of attention, one has plausible fitness functions, but the mutation model lacks a microscopic underpinning.

Genotypes: In the most straightforward picture, genotypes are identified with linear arrangements of L sites. Each site i is equipped with a variable σ_i which may take values from a set V_i. This way, a configuration may be denoted by $\boldsymbol{\sigma} \in V_1 \times \ldots \times V_L$. In the classical context, 'sites' are identified with '(gene) loci', and 'variables' σ_i with 'alleles'; then, $\boldsymbol{\sigma}$ is the configuration of a so-called *multilocus system*. The V_i may be very 'large' sets (comprising all possible alleles at a locus), but sometimes the simple lumping into *wildtype* (+) and *mutant* (−) alleles is sufficient, i.e. $V_i \equiv \{+,-\}$. The corresponding genotype space, $\{+,-\}^L$, was introduced and visualized by Wright (1932) [174].

In the molecular context, $V_i \equiv V$ may be the nucleotide alphabet {A(denin), G(uanin), C(ytosin), T(hymin)}; however, a binary alphabet ($V = \{0, 1\}$ or $V = \{+,-\}$) is often used instead, where the variables are lumped into purins (A,G) and pyrimidins (C,T). In both cases, configurations may be interpreted as DNA sequences (or RNA sequences if the letter T is substituted by U(racil)). Consequently, V^L is known as *sequence space*.

While the classical picture was the primary one historically, it may be considered as an *effective theory* today – like some kind of Landau-Ginzburg-Wilson theory in physics. For the purpose of this review, we shall move freely between the (molecular) sequence and the (classical) multilocus pictures. In both cases, a continuous limit is often appropriate, although genetic information is discrete in principle. This may be achieved through $L \to \infty$ (the *infinite sites model* [96], see also [41]), or by choosing V as infinite or even continuous.

In certain classical contexts, the 'genotype' is an indirect construct. Rather than being defined mechanistically as a sequence of letters, it is defined through its effect on the phenotype, as will become clear in a moment.

Phenotypes: We will single out two phenotype spaces, which are representatives of the molecular and the classical pictures, respectively.

In the molecular picture, if taken seriously, it is clear right from the beginning that the genotype-phenotype mapping is vastly complicated. No matter what is considered as the phenotype, it will certainly include several levels of organization, the first of which is *protein folding* – one of today's big unsolved problems. One useful compromise is, therefore, to consider genes which code for RNA (as opposed to proteins), or RNA molecules as such, populations of which may be replicated in the lab [15, 16]. Much of an RNA molecule's properties is determined by its secondary structure, which, therefore, serves as a legitimate phenotype. Unlike DNA, which consists of two complementary, base-paired molecules, RNA is single-stranded and partly folds back upon itself, forming *stems* where base pairings occur, and *loops* where letters are unpaired. The secondary structure is therefore determined by the collection of base pairings in a molecule, i.e. it is a *discrete* quantity. It is computationally far more accessible than protein structure. The prediction of the minimum free energy structure may be achieved with the help of algorithms based on dynamic programming [180, 77]. A variety of algorithms and different sets of thermodynamic parameters have been used. Although the details of the results are highly sensitive to the particular choices, several qualitative features and many statistical properties (i.e. the frequency distribution of structures) seem to be largely independent of the

prediction method, see [162] and references therein. Exhaustive enumeration has been performed for all RNA sequences with $L = 30$ [66, 67]. This reveals that the frequency distribution of structures follows a generalized version of Zipf's law [66]. With increasing length, there are only few common structures and many rare ones, so that only a few phenotypes will matter in practice. Sequences folding into these common structures percolate sequence space. On the other hand, sequences folding into almost all common structures can be found within a small distance of any random sequence [67].

Although it may be debated whether the phenotype is representative, it is felt that it captures some typical features of biological macromolecules, like long-range, asymmetric and irregular interactions, and a many-to-one mapping from genotype to phenotype.

In the classical picture, the phenotype space is a *trait space*. Most traits are so-called *quantitative traits*, i.e. they either vary continuously (e.g. body height or milk yield), or there is a large number of possible values that may be adopted (e.g. the number of bristles on the abdomen of *Drosophila*). This is opposed to discrete traits (peas may be either green or yellow). For practical reasons, one is often restricted to one or a few traits. The mapping from genotype to phenotype relies on the assumption that a large number of loci contributes to a given trait z. In the simplest case, they act independently, so that z may be written as $z = \sum_j \alpha_j \sigma_j$, where α_j is known as the *effect* of site j on the trait, and $\boldsymbol{\sigma} \in \{0, 1\}^L$. Similarly, $z_i = \sum_j \alpha_{ij} \sigma_j$ if $\boldsymbol{z} \in \mathbb{R}^n$. Much of the quantitative genetics literature assumes $\alpha_j \equiv \alpha$, but loci of major and minor effects have been identified recently (for review, see [127]), which reveals the relevance of inhomogeneities across sites. However, even the homogeneous models have been extremely successful in describing and predicting breeding experiments, i.e. short-term (artificial) evolution.

Genotypes from phenotypes: Owing to the ready observability of the phenotype and the virtual inaccessibility of the genotype in many situations, the phenotype is often considered as the primary quantity, whereas the genotype is a derived construct. A prominent example is quantitative genetics. For a trait vector $\boldsymbol{z}$, $\boldsymbol{x}$ denotes the corresponding genetic contribution. In general, a trait is determined by both genetic and environmental contributions. In line with the fitness landscape picture, we ignore environmental effects for the purpose of this review, thus reducing quantitative genetics to the case $\boldsymbol{x} = \boldsymbol{z}$. We are well aware that this comes close to castration of an important subject; however, this restricted view will suffice to understand those mutation models and fitness landscapes from quantitative genetics which we will meet in the context of mutational degradation.

We have, so far, defined genotypes and phenotypes. Building on this, we shall now introduce mutation models (which act on the genotype) and fitness landscapes (mappings from genotypes into fitness values, ideally via the phenotype). In doing so, we shall classify the material according to logical and/or mathematical aspects. This way, models of molecular evolution may appear next to those used in animal breeding. This should not be misunderstood as neglect of the historical context or the biological motivation; however, it is felt that the classical and molecular fields should (and do!) intermingle, and much can be gained by considering their mutual relationships.

4.1 Mutation models

If the genotype is a collection of sites, it is usually assumed that all sites mutate independently and experience the same transition probabilities. With binary variables at the sites, mutation is either chosen symmetric or unidirectional. Symmetric mutation is more adequate for the molecular context, whereas unidirectional mutation is often used in the classical regime. The notion behind the latter is that, actually, multiple alleles per site are assumed, but they are lumped into a (small) 'wildtype' and a (large) 'mutant' class, where mutations from wildtype to mutant are predominant and back mutations negligible, due to sheer entropic reasons.

With $\boldsymbol{\sigma} \in \{+,-\}^L$ and symmetric mutation with probability p per site at every reproduction event, the mutation probability from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma'}$ reads

$$v_{\boldsymbol{\sigma'}\boldsymbol{\sigma}} = p^{d(\boldsymbol{\sigma'},\boldsymbol{\sigma})}(1-p)^{L-d(\boldsymbol{\sigma'},\boldsymbol{\sigma})}\,, \tag{17}$$

where $d(\boldsymbol{\sigma'},\boldsymbol{\sigma})$ is the *Hamming distance* of sequences $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$, i.e. the number of sites where $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$ differ. We shall, in what follows, refer to (17) as *sequence space mutation*, since it is mainly used in this context [38].

Constant mutation probabilities over sites clearly represent an idealization which is seldom realistic. The existence of *mutational hot spots* is very well documented in molecular evolution, for review see [160] or [113, pp. 74–77]. However, many important aspects are expected to be captured by the homogeneous model already.

Owing to the more abstract nature of the genotype, the mutation models of quantitative genetics are far less mechanistic. For genotypes $\boldsymbol{x}$ from a (possibly continuous) state space $\boldsymbol{X}$, let $f(\boldsymbol{x'},\boldsymbol{x})$ be the *mutation distribution*, i.e. the probability density for a mutation from $\boldsymbol{x}$ to $\boldsymbol{x'}$ conditional on the assumption of a mutational event ($\int_{\boldsymbol{X}} f(\boldsymbol{x'},\boldsymbol{x})d\boldsymbol{x'} = 1$). It is often assumed that $f(\boldsymbol{x'},\boldsymbol{x}) = g(\boldsymbol{x'}-\boldsymbol{x})$; this is a generalization of the *random-walk mutation model* as introduced by Crow and Kimura [26]. One favourite choice for g is a multivariate Gaussian distribution, as used by Lande (1976) [109]. This is the only quantitative genetics mutation model which we will need in what follows; we will therefore leave it at that and refer the reader to textbooks and reviews, e.g. [12, 19, 118], for more variety.

Quite generally, the mutation model is less of a worry than the fitness landscape. The lack of knowledge about the latter is reflected by the jungle of choices which are in use, and which we shall now try and explore.

4.2 Fitness landscapes

We shall use the notion of fitness landscape in the way brought up by Kauffman and Levin [89]: As a mapping from genotype space into the real numbers. We shall be exclusively concerned with stagnant environments, i.e. the landscape is fixed. This is a severe restriction and excludes, for example, extinction events due to environmental catastrophes. On the other hand, it is often legitimate for short-term evolution and particularly so for evolution in the lab, where conditions may be kept constant over a relevant time scale when organisms with short generation times (like viruses) are used.

We shall proceed from the simple to the more 'complex'. The trivial landscape is all *flat*. It is the major theme of the aforementioned *neutral theory of molecular evolution* developed by Kimura in the sixties as a response to the fresh discovery of an overwhelming and unexpected amount of variation at the molecular level; see [97]. Since this was hard to explain under the then-standard multiplicative fitness function (see Eq. (18) below), the bold consequence was drawn that the vast majority of mutations is *selectively neutral*, i.e. has no effect on fitness; this gave birth to the neutral theory.

Although a flat fitness landscape does certainly not qualify as a model for evolution in general, the neutral model rightly serves as the basis for analysis of molecular data from certain genomic regions (like large parts of the mitochondrial DNA, which are non-coding and may be taken to be free of selection in a good approximation), as an important null hypothesis (which is actually often hard to reject!), and as a toy model which allows explicit calculation of many quantities of interest.

Some 'local' information about the fitness landscape may be gained from mutation accumulation experiments, where recombination is prevented with the help of genetic tricks, and selection is relaxed as far as possible. In such experiments (performed mainly with *Drosophila*), one measures fitness (in terms of viability, or number of offspring, or both) of the progeny of *a single* genotype over many generations, i.e. as a function of the *average number of mutation events*. From the observed decrease in fitness, the deleterious mutation rate is estimated and can be as high as one mutation per genome per generation for detrimental effects of a few percent, and lethal mutations much rarer (see [130, 136, 28, 81, 151, 24, 86, 118], [94, 95] for a contrasting view, and [106] for a review). Since the mutant genotypes are not accessible, the information gained does not allow a reconstruction of the fitness landscape proper, not even locally. The simplest landscape *compatible* with the observations is one with a unique fittest genotype, the wildtype $++++\ldots++$; every '$-$' site corresponds to one deleterious mutation with detrimental effect s each. If deleterious effects act independently across sites, an individual carrying j mutations has Wrightian fitness

$$w_j = (1-s)^j\,; \tag{18}$$

this is the so-called *multiplicative fitness function*. The corresponding Malthusian fitness is linear, $r_j = -\alpha j$, where $\alpha = -\ln(1-s) > 0$ in due course[d]; this is often associated with a *Mount Fujiyama landscape* [92]. In both cases, there is no interaction between sites – this is known as lack of *epistasis*.

These fitness functions are examples of what we would like to call *permutation invariant* landscapes, meaning that the fitness of a configuration is invariant under permutation of sites. If sites are equivalent for the mutation model as well, a drastic reduction of relevant dimensions ensues. For $V = \{+, -\}$, to which we will adhere in what follows, the number of '$-$' sites in a configuration serves as a valid description for most purposes (with the exception of some aspects of Muller's ratchet, see below), and the dynamical equations simplify considerably. The fitness optimum is at the boundary of the (now one-dimensional) space; one speaks of *directional selection*.

Epistasis, in its simplest form, is inherent in quadratic fitness functions,

$$r_j = -(\alpha j + (\gamma/L) j^2)\,, \tag{19}$$

cf. [100, 22]. If this function is monotonic with its maximum at the boundary, one still has directional selection, but, depending on whether fitness is a concave ($\gamma > 0$) or convex ($\gamma < 0$) function, existing mutations have an aggravating or an alleviating effect on further ones, which is termed *synergistic* (or *positive*) and *diminishing returns* (or *negative*) epistasis, respectively [100, 22]. The mutation accumulation data seem to reveal some synergistic epistasis; see the discussion in [100].

An extreme form of epistasis (which changes from positive to negative) is *truncation selection*, where fitness is a step function of the number of deleterious mutations, i.e. for some k,

$$w_j = \begin{cases} 1 & \text{for } j \le k \\ 1-s & \text{for } j > k\,. \end{cases} \tag{20}$$

Again, an extreme case of this is what was originally called the *single-peaked landscape* (SPL). Here, $k = 0$, i.e. only one configuration in the space (the 'wildtype') has a selective advantage, whereas all others (the 'mutants') are equally unfit. Following recent usage, we rechristen it *sharply-peaked landscape* in agreement with the previous abbreviation. The SPL was originally suggested as a model for prebiotic evolution [38]. Even here, it should not be considered as more than a toy model. It may well describe certain restricted regions of the genome, for example the few sites in the active centre of an enzyme, the function of which is likely to be destroyed by almost any mutation which comes along. On the other hand, a mutation which hits the beta sheet in the body of a protein may go unnoticed – this is an aspect of *neutrality*. In general, it is obvious that fitness landscapes should include *compensatory mutations*, i.e. instances where a mutation at a second site undoes the harm done by the first. In the SPL, there is no chance for compensatory mutations, no matter what the mutation mechanism. Nice examples of compensatory mutations are present in

[d]Note that Malthusian fitness does not suffer from being negative.

the stem regions of RNA secondary structures. Since these distinguish only between paired and unpaired regions irrespective of the particular bases at the individual positions, base pairings may be destroyed by a point mutation, but may be re-established by a second, compensatory one, cf. [158, 75]. In general, the opportunity for compensatory mutations depends on the mutation model as well as the fitness landscape. In the multiplicative (18) and the quadratic (19) landscapes, for example, there are no compensatory mutations if mutation is unidirectional, but plenty of them when mutation is symmetric.

In the quadratic fitness function (19), one may interpret j as a quantitative trait z. The corresponding Wrightian fitness corresponds to a Gaussian distribution, which may be written as a function of the trait value: $w(z) = \exp(-s(z - z_{opt})^2)$, where z_{opt} is the optimal phenotype, and s determines the strength of selection[e]. If the optimum is in the interior instead of at the boundary, one has *stabilizing selection*. This is presumably a common form of selection on quantitative characters in the sense that selection favours intermediate trait values in preference to either extreme. In contrast to the situation with directional selection, the fittest genotype is not unique in the configuration picture. There is a large proportion of compensatory mutations, which restore the fitness without restoring the original genotype.

We will next consider more elaborate landscapes and, for that purpose, return to the configuration space $\{+, -\}^L$. Two questions will be crucial: How rugged is a landscape, and how large is its degree of neutrality?

Neutrality is easily defined as the average number of neutral neighbours per configuration, where 'neighbourhood' is defined with respect to the Hamming distance. Ruggedness is, intuitively, related to the abundance of local peaks, as well as the depths of the valleys separating them. Experimentally, the ruggedness of landscapes has been explored in the multilocus context with the help of special crossing experiments; for review see Whitlock et al. (1995) [171]. The authors stress the abundance of nonlinear interactions, with special emphasis on those leading to multiple fitness peaks. A principal difficulty prevails, however: Since it is impossible to explore all dimensions of the genotype space, it is impossible to decide whether high spots are isolated, or whether they are connected by ridges in higher dimensions. One must therefore be satisfied to know that landscapes are rugged, without knowing whether they are peaked. In line with this, it has been argued on theoretical grounds that, in higher dimensions, isolated fitness peaks are extremely rare. High points are likely to be connected by a 'bypass' in at least one of the many dimensions [56, 57].

Formally, correlation functions and the density of local optima have been proposed as measures of ruggedness; see, e.g., [92, 156, 150]. Since the exhaustive exploration of a landscape is only possible if the number of sites is small, one has to resort to some kind of statistical evaluation. This may be done with the help of *adaptive walks* [92, 119].

[e]We use s as a general variable to indicate the strength of selection. It may express a selective advantage or a disadvantage depending on the context.

An adaptive walk proceeds via random choice of a neighbour with higher fitness in every step, until no higher fitness is found; i.e. it stops at a local peak. The distribution of the lengths of these walks, as well as the properties of endpoints reached (e.g. whether they are isolated peaks or ridges), may be used to characterize a given landscape. Likewise, *neutral walks* attempt to find a neutral neighbour in each step so that the distance from the starting point increases [66]. The distribution of their lengths characterizes the neutral properties of landscapes.

We shall now consider three representative families of fitness landscapes in some more detail.

RNA, neutral networks, and holey landscapes: Since, within the RNA world, fitness is a function of the (secondary) structure, an equivalent of our permutation-invariant landscapes may be obtained by distinguishing one 'target' structure and assigning selective disadvantages to structures according to some suitably defined distance from the target [85]. A corresponding SPL then results from the assumption that all but the target structure are equally unfit [148]. However, assigning fitnesses to genotypes requires computation of the structure, which involves an enormous computational effort. Hence there is a large demand for simpler toy models which mimick the essential features. The pre-image of a given structure, its *neutral space*, has been modelled as the vertex set of a certain random graph, with an edge between vertices if they are neighbours in sequence space. This way, one random graph (termed *neutral network*) is associated with every structure [149]. A closely related concept is that of *holey landscapes*, where a fraction of randomly chosen genotypes is lethal [56, 57]. The crucial parameter in both cases is the mean fraction of neutral neighbours of a sequence. If it surpasses a critical value, giant components of neutral networks, or fit genotypes, appear and percolate sequence space; this is as observed for the more common RNA structures. The percolation threshold decreases dramatically with the number of sites involved, i.e. the dimensionality of the space [149, 56].

NK landscapes, spin glasses, and Hopfield Hamiltonians: It is generally believed now that landscapes are rugged due to the many interactions between genes and within genes; but there is no agreement how rugged exactly. A family of *tunably rugged* landscapes has been put forward by Kauffman and Levin [89] and further exploited by Kauffman and coworkers; for review, see [88, 92, 93]. These so-called NK landscapes owe their name to the assumption that there are N (in our notation: L) sites, each of which interacts with K other sites in a random manner, as originally motivated by metabolic [90] gene regulatory [91] networks. These landscapes correspond to spin glasses with $\tilde{K} = K + 1$ interacting spins, i.e.

$$r_{\boldsymbol{\sigma}} = \sum_{\{i_1,\dots,i_{\tilde{K}}\}} J_{i_1 \dots i_{\tilde{K}}} \sigma_{i_1} \dots \sigma_{i_{\tilde{K}}} \,, \tag{21}$$

where the $J_{i_1 \dots i_{\tilde{K}}}$ are independent, identically distributed (i.i.d.) random variables for each different set of indices $\{i_1, \dots, i_{\tilde{K}}\}$. They can take positive as well as negative values,

thus causing *frustration*. The landscape is tunably rugged through $\tilde{K}$. For $\tilde{K} = 1$, one has an (inhomogeneous) Fujiyama model, i.e. one unique peak and a highly correlated landscape; it is also known as *random field paramagnet* [145]. For $\tilde{K} = 2$, one has the well-known Sherrington-Kirkpatrick spinglass (for review, see [126]); it was first suggested in the evolutionary context by Anderson [5]. It is rather correlated but has a relatively large number of local optima. One special choice for the interaction constants is $J_{ij} = \sum_{p=1}^{P} \xi_i^p \xi_j^p$. The resulting energy function is known as *Hopfield's Hamiltonian* and was originally designed to describe the operation of *neural* (as opposed to *neuTral!*) *networks*. Here, the $\boldsymbol{\xi}^p := (\xi_1^p, \dots, \xi_L^p)$, $p = 1, \dots, P$, are configurations randomly chosen from V^L. They represent the P *patterns* learnt by, and stored in, a neural network.

Hopfields's Hamiltonian was used as a fitness landscape by Leuthäusser [112] and Tarazona [164]. It may be rewritten as

$$r_{\boldsymbol{\sigma}} = \frac{1}{L} \sum_{p=1}^{P} \big(\sum_j \xi_j^p \sigma_j \big)^2 \,. \tag{22}$$

Here the ruggedness is determined by P, in a statistical sense; of course it also depends on the particular choice of the $\boldsymbol{\xi}^p$ in every single case. For $P = 1$, one has the so-called *Mattis Hamiltonian*. Noting that $\boldsymbol{\xi}^1$ may be chosen as '$+ + + + + \dots +$' without loss of generality, it is clear that the Mattis Hamiltonian corresponds to the permutation-invariant quadratic fitness function (19) with $\alpha = 0$ and $\gamma = 1$.

For $\tilde{K} = N$, one has what is called the random energy model [29] in the terminology of spin glasses, i.e. a fully uncorrelated landscape with a random fitness value for every genotype. However, as a fitness landscape, it is older and has its roots in classical population genetics. When thinking about the effects of mutation, Kingman (1977) stated 'the tendency for most mutations to be selectively disadvantageous, presumably because they upset the evolutionary house of cards, built up by the careful improvement of many generations .... This suggests a model ... in which mutation always results in a completely novel allele, and in which the fitness of the mutant is chosen from a fixed fitness distribution' [101]. This became known as the *house-of-cards model*. In this formulation, the model is annealed whereas evolution on proper fitness landscapes is a quenched problem, i.e. the interaction constants are frozen during the evolutionary process. However, for reasons of storage economy, the random energy model is usually simulated as annealed in the evolutionary context (e.g. [4]), which does not seem to produce noticeable artifacts.

For $\tilde{K}$ increasing from 2 to N, there is an increasing degree of conflicting constraints, as might be typical of biological evolution, and 'frustration' becomes more and more abundant. The NK model has been extensively studied in the context of adaptive walks; for review, see [92].

The hierarchy of spin glass models seems straightforward and plausible. It should be noted, however, that the present state of knowledge is fairly incomplete. Little is known rigorously, not even about the energy functions as such (cf. [126]), let alone the evolutionary dynamics on these landscapes.

One toy model which is between the smooth quadratic landscapes and the rugged spin glasses is *Onsager's landscape* [8] which is related to (21) through $\tilde{K} = 2$ and $J_{ij} = \delta_{j,i+1}$. Here, fitness is determined by the number of *domain walls* in a sequence. As a consequence, there are compensatory mutations, flat ridges, as well as a high degree of neutrality. Of course, the symmetries are artificial, but, as with two-dimensional Ising models with nearest neighbour interaction, they make an exact solution possible (see Sect. 5.1.2 below).

Multiple quantitative traits: Spin glass models do not have, and do not suggest, a phenotype, which is unsatisfactory from the biological point of view. This is different for the multiple quantitative traits (MQT) model which has been put forward very recently [72]. There are L sites and T traits, with every trait influenced by $K \leq L$ randomly chosen sites. For sequences as bit strings, $\boldsymbol{\sigma} \in \{0,1\}^L$, traits are modelled as $z_i = \sum_{j=1}^{L} a_{ij}\sigma_j$, $i = 1, \ldots, T$, with $a_{ij} = 1$ if site j affects trait i, and $a_{ij} = 0$ otherwise. Wrightian fitness is based on the sum of quadratic deviations of the traits from their randomly chosen optimal values, i.e.

$$w(\boldsymbol{z}) = \exp(-s \sum_{i=1}^{T} (z_i - z_i^{opt})^2), \tag{23}$$

where s again measures the strength of selection.

The resulting fitness landscape is tunably rugged through the product KT. As long as $KT \ll L$, the sets of loci linked to the different traits rarely overlap; as a consequence, traits evolve independently, and genotypes are possible which simultaneously optimize all traits. If $KT \gg L$, on the other hand, most loci will affect more than one trait, and there is no 'perfect' genotype [72].

The main difference between MQT and spin glass models is that, in MQT landscapes, there are clusters of high-fitness genotypes instead of maxima randomly distributed throughout the space. This may be seen as a higher-dimensional issue of directional versus stabilizing selection. In the former case, maxima are unique, and the degree of neutrality increases with distance from them; in the latter case, maxima tend to be plateaus with some degree of neutrality.

We have, so far, met a whole zoo of fitness landscapes, which would benefit from a general characterization and classification. An attempt in this direction has been made recently through the concept of *additive random landscapes* [150, 155]. This is a large class of fitness landscapes both tunably rugged and tunably neutral, which includes spin glass models, among others.

4.3 Scaling and limits

A very important issue for both mutation models and fitness landscapes is that of *scaling* as a function of the system size. For both of them, extensive and intensive scalings are

in use. *Extensive* quantities scale linearly with L, whereas *intensive* ones are independent of the number of sites. For the mutation model, the distinction is straightforward on biological grounds. Extensive scaling is assumed with sequence space mutation (17), and motivated by the molecular picture: Every site has a fixed probability to mutate, and thus the mutation probability for the whole genome scales with genome size. In other cases, the genome (or the sequence, or the collection of sites) is taken as the primary quantity, and with it associated a fixed mutation rate U per year or generation, which corresponds to intensive scaling. In this context, one relies on the *infinite sites limit*, i.e. $L \to \infty$ under $L \cdot p = U \equiv$ const. Often, however, population size must be considered, too. In what follows, distinguishing carefully between various limits and scalings will be crucial to an understanding of apparently contradictory results. In particular, we shall meet the

$$\textit{infinite genome limit:} \quad L \to \infty \quad \text{under } L \cdot p = \text{const}, \quad \text{followed by } N \to \infty\,, \tag{24}$$

and the

$$\textit{infinite population limit:} \quad N \to \infty \quad \text{followed by } L \to \infty \quad \text{under } p = \text{const}. \tag{25}$$

Note that we adhere to the term *infinite sites limit* in the above sense if nothing is implied about population size.

Quite generally, for the real situation with finite N and finite L but N smaller than the number of possible configurations, the population will move as a (relatively small) cluster in the configuration space and eventually sample all of space, but this will take extremely long. This is the motivation for the infinite genome limit. A side effect is, however, that every site experiences at most one mutation event ever. For example, if $++++++$ has mutated to $+-++++$ in one individual and to $+++-++$ in another, the configuration $+-+-++$ can *never* be accessed in *any other* individual. Consequently, the system cannot be ergodic, and the initial state of a population is important even for infinite times. Deviations from such a model are apparent in human mitochondrial sequence data, e.g. [170].

The infinite population limit, on the other hand, relies on the ergodicity of the real system. It corresponds to the thermodynamic limit in statistical physics, with L the number of particles and N the number of copies of the system. The calculated quantities are to be understood in the sense of time averages. It is particularly adequate when the effective configuration space is drastically reduced, e.g. due to an abundance of lethal mutations. The side effect here is that the averaging may take longer than evolutionary (or Monte Carlo) time.

In summary, both limits may have their problems, which should be kept in mind when interpreting the results. In particular, results obtained with one of them may not be used to draw conclusions about the other; this seems to have been overlooked in places. For example, Derrida and Peliti [31] argue on the grounds of (24) about the quasispecies model (to be described in Sect. 5.1 below), which relies on the limit (25). This leads to results at variance with the properties of the original model.

Let us now turn to scaling of fitness landscapes. For Malthusian fitness, extensive scaling implies that $r_{max} - r_{min}$ increases linearly with L. For Wrightian fitness, this corresponds to $w_{max}/w_{min} \to \infty$ and, with the normalization $w_{max} = 1$, to $w_{min} \to 0$ for $L \to \infty$, that is, genotypes become more and more lethal. With intensive scaling, on the other hand, w_{min} is bounded away from zero, and there are no lethal genotypes.

The SPL (sharply-peaked landscape) with $s < 1$ is an example of intensively-scaled landscape (see, however, [46]), whereas the multiplicative fitness function is extensive.

Both types of scaling are used in both the molecular and the classical contexts, often without awareness of the distinction. In particular, all combinations of mutation and fitness scalings show up: all-extensive, all-intensive, and mixed. It should be noted that the all-extensive and the all-intensive scalings are related to each other through a corresponding scaling of (continuous) time, i.e. dt is replaced by Ldt, see the discussion in [169]. This is often overlooked since only few investigations consider dynamical aspects at all. In any case, it should be kept in mind that any real sequence is finite, and $L \to \infty$ is just a way of making life simpler — a good one though, since in these applications, 100 is often closer to infinity than to one. When interpreting the results, attention should be paid to correct adjustment of parameters to meet the finite size context.

5 Modelling mutational degeneration

We have, so far, worked hard on the list of ingredients. As a reward, we are now in a position to piece together mutation-selection models from the elements of our construction set.

A large body of work on such models has actually been related to mutational degradation in asexual populations. These models predict upper limits for the mutation rates above which mutation can no longer be controlled by selection, the most important phenomena being error thresholds, Muller's ratchet, and mutational meltdowns.

At first sight, the kind of question seems ill-posed. Wouldn't the primary effect of mutation be to introduce new variation on which selection can act to bring about evolutionary progress? According to a commonly-held view, however, today's populations have arrived at a state of elaboration where most mutations are deleterious. Whereas this is a 'finalistic' point of view which one need not necessarily share, the argument is indisputable for two classes of problems:

1) Consider a sexual population which, with the help of the larger adaptation potential of recombination, has arrived at a fitness peak. Let then a clonal lineage split off. Here, *clonal* means that the genome is passed on unaltered from parent to offspring (apart from mutations). This may occur through parthenogenesis (in animals) or vegetative reproduction (in plants). Actually clonal lineages arise frequently in both the animal and the plant kingdom [117], even in fairly 'high' organisms (e.g. certain lizards).

In the absence of recombination, such populations will incessantly slide off the fitness peak due to deleterious mutations and genetic drift. This process is known as *Muller's ratchet*. It may lead to the extinction of populations through so-called *mutational meltdowns* which provide one favoured explanation for the observed short life of clonal lineages [117].

2) In the context of prebiotic evolution, and the evolution of viruses and bacteria which do not have recombination (however, some viruses and bacteria do recombine, see [157, 124, 123]), the mutation-selection equilibrium is of primary importance. With respect to the highest peak in the landscape, every mutation is either neutral or deleterious. Their joint effect on the equilibrium fitness is known as the *mutation load*, $\ell := r_{\max} - \bar{r}$ (recall that $\bar{r}$ is the mean (Malthusian) fitness of the population). This is important in the context of the evolution of mutation (as well as recombination) rates. After all, the evolution of repair mechanisms along with that of the eukaryotic cell suggests that high mutation rates are not unconditionally desirable. Actually, certain models predict a complete loss of genetic structure when the mutation rate surpasses a critical value; this is known as the *error threshold*.

We shall now embark on these phenomena, starting with error thresholds (they build on mutation and selection only); we shall proceed with Muller's ratchet, which requires drift as well. Finally, mutational meltdowns will be considered, which additionally require an explicit model of population dynamics. In contrast to common usage, we are reluctant to speak of 'models of error thresholds' or 'models of Muller's ratchet'. Instead, we prefer to think about models in terms of their ingredients, and to examine which features they exhibit in dependence of these ingredients.

As we shall see, clear definitions of the phenomena to be examined are lacking in places. Things are clearest for mutational meltdowns but less so for error thresholds. The phenomena may, however, be fairly well *described*, if not defined, in terms of prototype models which exhibit the respective behaviour. These are, not surprisingly, the historical archetypes, and they all employ the permutation invariant class of fitness landscapes.

5.1 Error thresholds

They are the most prominent features of the so-called *sequence space models*, as inspired by the discovery of the molecular structure of genes, and originally aimed at prebiotic evolution in the RNA world. The most well-known sequence space model is Eigen's *quasispecies model* [38] which may be understood as the coupled mutation-selection model (8), but with alleles replaced by sequences $\boldsymbol{\sigma} \in \{+,-\}^L$, and symmetric mutation according to Eq. (17). That is, the ODE system reads

$$\dot{p}_{\boldsymbol{\sigma}} = \sum_{\boldsymbol{\sigma}'} v_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} r_{\boldsymbol{\sigma}'} p_{\boldsymbol{\sigma}'} - \big(\sum_{\boldsymbol{\sigma}'} r_{\boldsymbol{\sigma}'} p_{\boldsymbol{\sigma}'}\big) p_{\boldsymbol{\sigma}} \tag{26}$$

for the relative frequencies of sequences $\boldsymbol{\sigma}$.

5.1.1 The prototype model

Whereas sequence space mutation is a natural mutation model for the situation considered, there is no canonical fitness landscape. The toy model which has received a lot of attention is the SPL (sharply-peaked landscape) with selective advantage s of the wild type. (The quasispecies community has created its own terminology, for instance, this favourable sequence is called 'master sequence'; we shall, however, avoid using redundant terminology here.)

This toy model is so well-known because its stationary state exhibits a behaviour reminiscent of a phase transition. Whereas the population is closely centred around the favourable sequence at small mutation rates, it is close to evenly distributed over the space when p surpasses a critical value. This is shown in Fig. 3 for the closely related paramuse model, to be described in the next subsection. This may be interpreted as mutation becoming so strong that it can no longer be counteracted by selection, which, in due course, leads to the (effective) loss of the favourable sequence. This feature became known as the *error threshold*.

The quasispecies literature up to 1989, and the sharply-peaked landscape in particular, has been comprehensively reviewed [39], wherefore we shall not dwell on details here. It is, however, important to note that, in spite of the apparent simplicity of this landscape, there is no exact analytic solution known, not even for the stationary state. One relies on numerical solutions, either through integration of the ODE system, or determination of the dominant eigenvector. Both take advantage from a reduced representation of the mutation-reproduction matrix, which is available thanks to the symmetries of the system. However, the special SPL situation is readily approximated by neglecting back mutation from the mutants to the favourable sequence, which are very rare events indeed. This yields the critical mutation rate [39]

$$p \simeq 1 - \left(\frac{1}{1+s}\right)^{\frac{1}{L}} \simeq \frac{s}{L} \tag{27}$$

to first order in s, in good agreement with the numerical value. It should be noted, however, that a different approximation technique as proposed in [3], which treats the sites as statistically independent entities, leads to severe artifacts. For example, it predicts a maximum s above which the error threshold does not occur, no matter how large the mutation rate; such a feature is absent from the original system.

Relation (27) is readily interpreted as a maximum mutation probability allowed at a given selective advantage and sequence length, or, alternatively, in terms of a maximum sequence length that may be correctly maintained under a given selective advantage and mutation probability. Both may be taken as indicative of the need for the evolution of repair mechanisms in the course of the evolution of the larger eukaryotic genomes [125]. In invoking such arguments, however, it is usually overlooked that the inverse relationship (27) relies on the toy model of a sharply-peaked landscape as generalized to whole genomes, which is, at best, dubious (see Sect. 4.2); this point is discussed in detail

in [172]. It should be added that, although we have not yet dealt with finite populations in sequence space, even very large populations (viruses: $N \simeq 10^{12}$; *Drosophila*: $N \simeq 10^{6}$) would be very unlikely to find the isolated peak, even if sequence lengths were tiny (say $L = 1000$).

On the other hand, the entire error threshold phenomenon has occasionally been dismissed on the grounds that it its based on an SPL, e.g. [22], which is premature as long as it is unknown which other landscapes exhibit this phenomenon, too. Both lines of argument make clear that it is now imperative to study other fitness landscapes. Since there are only few clues as to what a biologically relevant fitness landscape is, it is imperative to study a variety of choices. But this is paved with obstacles (recall that, even for the SPL, there is no exact analytical solution). The only benevolent case is the Fujiyama (resp. multiplicative) landscape (18), where the independence of the sites allows for a straightforward solution. The stationary state has been given by several authors [153, 73, 8]. Namely, the frequency p_j of sequences with j '$-$' sites follows a binomial distribution,

$$p_j = \binom{L}{j} a^j (1-a)^{L-j}\,, \tag{28}$$

where the parameter a depends smoothly on the relative strength of mutation and selection; in particular, $a = 0$ for $p = 0$, and $a = 1/2$ for $p = 1/2$. Clearly, there is *no error threshold*; instead, both the mean fitness and the genetic structure of the population fade away gradually with increasing mutation rate. For all other landscapes, more elaborate approaches are required.

We shall therefore proceed by discussing methods of analysis. It will then be necessary to clarify more precisely what an error threshold is supposed to be. After this excursion, we shall summarize the results that have been obtained. We shall finally discuss the experimental clues concerning the error threshold phenomenon.

5.1.2 Methods of analysis

The approaches used are all from statistical physics; the most important ones involve Ising models. Leuthäusser [111, 112] established an exact equivalence between a discrete-time version of the quasispecies model, and a 2D classical Ising model. To see this equivalence, consider a 2D square lattice anisotropic Ising system with Hamiltonian

$$\mathcal{H}_{class} = \sum_{i=1}^{n} \Big(E(\boldsymbol{\sigma}^i) - \sum_{j=1}^{L} J \sigma_j^{i+1} \sigma_j^i \Big) \tag{29}$$

with $\boldsymbol{\sigma}^i$ the spin configuration of the i'th row, $E(\boldsymbol{\sigma}^i)$ its energy, and J an interaction constant. The corresponding row-to-row transfer matrix (cf. [165]) has elements

$$\mathcal{T}_{\boldsymbol{\sigma}'\boldsymbol{\sigma}} = \exp\big(-\beta E(\boldsymbol{\sigma})\big) \exp\big(\beta J \sum_j \sigma'_j \sigma_j\big)\,, \tag{30}$$

where β is the inverse temperature. In close analogy, the elements of the mutation-reproduction matrix, $\mathcal{T} := \mathcal{VW}$ of Eq. (2), may be written as

$$\mathcal{T}_{\boldsymbol{\sigma}'\boldsymbol{\sigma}} = (p(1-p))^{L/2} \cdot \exp(r_{\boldsymbol{\sigma}}) \exp\left(\beta J \sum_j \sigma'_j \sigma_j\right), \tag{31}$$

where $\beta = -\ln \frac{p}{1-p}$ and $J = 1/2$; note that $(p(1-p))^{L/2}$ is a constant factor independent of the spin configuration.

The Ising system involved here is anisotropic: It has nearest-neighbour interaction *between* the rows (which correspond to mutation), but *arbitrary* interactions within the rows (the within-row interaction energy corresponds to the fitness of the configuration, the details reflecting the fitness landscape); see Fig. 4.

With this in mind, solving the evolution model is paramount to diagonalizing the transfer matrix of the corresponding Ising model. In particular, knowledge of the largest eigenvalue allows for the calculation of the stationary state. This equivalence was exploited in a number of applications. Tarazona [164] tackled the quadratic (Mattis) landscape (19), as well as Hopfield Hamiltonians (22); Franz and Peliti [46] and Franz, Peliti and Sellitto [47] examined the random energy model. On the whole, however, results have been surprisingly sparse. The reason seems to be that transfer matrices are hard to treat, not just due to their size, but due to the characteristic anisotropy of the interactions involved.

A related analogy which was described recently [8] circumvents this problem. It starts from the *paramuse model*, which had previously been suggested as an alternative to the quasispecies model [7]. It is the *para*llel *mu*tation-*se*lection model (32) as adapted to sequence space,

$$\dot{p}_{\boldsymbol{\sigma}} = (r_{\boldsymbol{\sigma}} - \bar{r}) p_{\boldsymbol{\sigma}} + \sum_{\boldsymbol{\sigma}'} m_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} p_{\boldsymbol{\sigma}'} \,. \tag{32}$$

Here, the mutation rates simply read

$$m_{\boldsymbol{\sigma}'\boldsymbol{\sigma}} = \begin{cases} \mu, & d(\boldsymbol{\sigma}', \boldsymbol{\sigma}) = 1 \\ -L\mu, & \boldsymbol{\sigma}' = \boldsymbol{\sigma} \\ 0, & \text{otherwise,} \end{cases} \tag{33}$$

where μ is the mutation rate per site.

With this, one arrives at a mutation-reproduction matrix of the form $\mathcal{H} = \mathcal{M} + \mathcal{R}$ instead of $\mathcal{T} = \mathcal{VW}$, which is exactly equivalent to the Hamiltonian of an *Ising quantum chain*. The limit which relates Ising quantum chains to their classical counterparts (cf. [104]) is just the short-generation limit which we have met in Eq. (11).

Explicitly, the quantum chain Hamiltonian reads (up to a constant term):

$$\mathcal{H} = \mathcal{M} + \mathcal{R} = \mu \sum_{j=1}^{L} \sigma_j^x + \sum_{j=1}^{L} \rho_j \sigma_j^z + \sum_{j,k=1}^{L} \rho_{j,k} \sigma_j^z \sigma_k^z + \ldots \text{terms up to } L\text{'th order}\,, \tag{34}$$

where, just for this moment, σ^x and σ^z denote Pauli's matrices, and

$$\sigma_j^a := \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \sigma^a \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1} \tag{35}$$

an L-fold tensor product with σ^a in the j'th place. Further, the collection of ρ's determines the fitness landscape; in most cases, only terms up to second order are involved (and difficult enough to handle).

As a note of caution, let us remark that, although the Hamiltonian is symmetric, the evolutionary dynamics does *not* have *detailed balance*. This is because the linear ODE, $\dot{\boldsymbol{x}} = \mathcal{H}\boldsymbol{x}$, has no stationary state at all; the normalized system (32) does have a stationary solution, but, due to the nonlinearities involved, it does not fulfil the conditions for detailed balance (i.e. $\partial f_{\boldsymbol{\sigma}}(\boldsymbol{p})/\partial p_{\boldsymbol{\sigma}'} \neq \partial f_{\boldsymbol{\sigma}'}(\boldsymbol{p})/\partial p_{\boldsymbol{\sigma}}$, where the $f_{\boldsymbol{\sigma}}(\boldsymbol{p})$ constitute the right-hand side of (32)).

The toolbox developed for quantum chains may then be applied to the evolution model. Some new techniques are, however, required to take care of the fact that it is not the quantum mechanical states which are relevant for the system; the problem remains one of classical probability [9]. So far, only two nontrivial cases have been worked out in detail, namely Onsager's landscape and quadratic fitness functions [169].

Another recent approach is the mapping of the evolution problem onto the Hamiltonian of directed polymers [55, 54]. Here, sequences are identified with elastic polymers, wandering in sequence space directed along the time axis, and subject to a potential. Here, mutation plays the role of elasticity, and the potential determines the fitness landscape; this leads to a transfer matrix very similar to (30). So far, the only application seems to concern the SPL, which corresponds to a *pinning potential* [54].

With all methods of analysis which we have mentioned, analytical (exact or approximate) studies are possible of order parameters and phase transitions, but numerical simulations are often required to resolve the population structure in detail. These, too, profit from the corresponding methods in statistical mechanics, see, e.g., [164].

5.1.3 Characterization of error thresholds

We have, up to now, carefully avoided to give a definition of the notion of 'error threshold'. Writing for a physical readership, we have tacitly interchanged it against the concept of *phase transition*, expecting the reader's approval. She or he may have noticed, however, that this need not, a priori, go together with the original descriptions.

Let us recall the original verbal description of mutational degradation. Mutation can no longer be counteracted by selection, so that genetic information is lost. This implies both a genetic and a fitness aspect, and let us add that there may also be an intermediate phenotypic aspect. But there is, as yet, no generally accepted definition of the error threshold phenomenon. Several descriptions are in use, but since error thresholds have been so closely tied to the prototype model (SPL with intensive scaling, and extensively-scaled symmetric mutation), criteria have been oriented towards this one, as well. The original criterion was the loss of the fittest sequence [38]; for the SPL (but not, necessarily, for other landscapes), this goes together with delocalization of the population over sequence space. Although pictures like Fig. 3 seem to speak a clear language, both properties are never met exactly with finite L for which the error threshold was originally described, but are expected to become exact only in the limit $L \to \infty$. In general, error thresholds appear to be strictly definable only in this limit. Mathematically speaking, this is because, for finite L and a primitive mutation matrix, there is a stationary distribution with $p_{\boldsymbol{\sigma}} > 0$ for all $\boldsymbol{\sigma}$; hence nothing can be lost, and no delocalization may occur. Physically speaking, no phase transition is possible with a finite number of sites. Note that the infinite population limit (25) is implied in the description by differential equations (26), and corresponds to the thermodynamic limit of statistical physics.

Let us therefore summarize possible error threshold criteria, with infinite sequences in mind. If the landscape has a single peak at sequence $\boldsymbol{\xi}$ say, the loss of the fittest sequence, and the vanishing of the average overlap ($u := \sum_{\boldsymbol{\sigma}} p_{\boldsymbol{\sigma}} \sum_i \xi_i \sigma_i$) with it, are the favoured criteria; it should be noted, however, that they need not give identical results [46, 172]. With this kind of landscape, $\boldsymbol{\xi} = +++++++\ldots+$ may be chosen without loss of generality; then, u corresponds to the magnetization of a classical spin system. To avoid confusion with the corresponding quantum mechanical quantity, we have previously termed it *surplus* [8]. If there are multiple (but still isolated) peaks, as in the Hopfield landscape (22) (they may stem from multiple patterns $\boldsymbol{\xi}^p$, as well as spin reversal symmetry), this order parameter performs one or several bifurcations, each indicating the 'loss of discrimination' between a pair of peaks [164]. If peaks are not unique, as with quantitative traits or RNA structures, the populations may spread over their neutral space, without being delocalized in phenotype space. Error thresholds are then reasonably defined as loss of the fittest phenotype; this is known as *phenotypic error threshold* [85, 148].

The delocalization criterion is appealing, but problematic. This is apparent from the stationary state (28) of the multiplicative landscape. Here, for every j, $p_j \to 0$ for $L \to \infty$ for any fixed but nonvanishing mutation rate. Hence, the distribution of genotypes is *de*localized over sequence space however tiny the mutation rate. In particular, the fittest sequence is lost. Nevertheless, one would not want to speak of an error threshold here (see above).

Another quibble with criteria based on knowledge of the fittest sequence is that they, in a sense, represent the standpoint of an omniscient observer. Due to the mixing of a genetic and a fitness aspect, such quantities are not observable independently of the fitness landscape. In contrast, the mean fitness of the population *is* observable, at least in principle, as is its genetic structure.

In line with our previous wording, we therefore propose phase transitions, or, alternatively, bifurcations of equilibria, as criteria for error thresholds. Actually, a general connection between bifurcations of equilibria and phase transitions (in the sense of non-analytic points of the free energy) has been conjectured, although this is far from being

proven rigorously [152, Ch. 5.7]. Both phase transitions and bifurcations go together with all threshold phenomena described so far, and those to be described in the sequel. On the other hand, they do not apply to the multiplicative landscape, as desired. One observation concerning the SPL may be interesting in this context. Although the above considerations are strictly applicable in the thermodynamic limit only, the situation here may be mimicked by a simple two-type model, where all unfit sequences are lumped together and mutation is *unidirectional* (no mutation back to the rare – fit – sequence) [10]. In this model, a (transcritical) bifurcation occurs with the mutation rate as the bifurcation parameter, in the course of which the fittest sequence is lost.

5.1.4 Results

Phase transitions in the usual sense may only occur if fitness and mutation both scale extensively, or both scale intensively, cf. [46]. Then, the result is a critical mutation rate per site, or per genome, respectively. If, on the other hand, fitness scales intensively and mutation extensively, one may obtain an inverse relationship between sequence length and mutation rate, as, for example, the one given in (27).

In the permutation-invariant quadratic landscape (19), a phase transition (of second order) is present if $\alpha = 0$, $\gamma < 0$ [8]. In this case, fitness is a convex function of the surplus, or, put differently, epistasis is negative. In contrast, lack of epistasis ($\gamma = 0$) precludes phase transitions; see our discussion of the Fujiyama (resp. multiplicative) landscape in the previous section. The same is true for positive epistasis ($\gamma > 0$) [22, 172]. These observations agree well with classical results predicting a higher mutation load in situations with negative epistasis as compared with positive epistasis [100]. With Onsager's landscape, one observes a second-order phase transition, too. In contrast to the quadratic landscape, where both surplus and mean fitness vanish at the critical mutation rate, the mean fitness continues to decrease beyond the nonanalyticity point in Onsager's landscape [8].

In Hopfield's landscape (22), only the vicinity of the highest peak is populated at small mutation rates. With increasing mutation rate, secondary maxima (which are less high but more abundant) take over, before, finally, the genetic structure is entirely lost. In line with this, there is not a single error threshold, but a sequence of bifurcations, each of which indicates the loss of discrimination between two patterns of the neural network [164]. Presumably, such a behaviour is typical of multi-peaked landscapes. The importance of such entropic effects can be mimicked by a toy model with one high and narrow, and a second less tall but broader fitness peak [154], the bifurcation structure of which may be analyzed exactly [10].

At the rugged end of the landscape zoo, the random energy model was investigated [47, 46]. Locally, i.e. in the vicinity of the highest peak, it was found to behave like an SPL; this is attributed to the complete lack of correlation.

Both Onsager's and Hopfield's landscapes may be considered as displaying (some suitable generalization of) negative epistasis. This is due to the fact that on average, with increasing distance from the maximum, a larger fraction of additional mutations either do no further harm, or even act in a compensatory manner. One might be tempted to conjecture that this type of epistasis is required for a phase transition to occur. It would be interesting to know the behaviour of the MQT (multiple quantitative traits) model (23), which may be interpreted as displaying something like *positive* epistasis (in the same way as quadratic landscapes with stabilizing selection do), but this has not yet been examined.

5.1.5 Error thresholds in finite populations

Unlike with infinite populations, we do not even attempt to define error thresholds for finite populations. Error thresholds are strictly definable only in the infinite population limit. However, obvious dramatic changes take place with finite populations as well, and we shall restrict ourselves here to describing instead of defining them.

In turning to finite populations, we leave the realm of difference or ordinary differential equations. Given a fitness landscape and a mutation model, the deterministic dynamics must be replaced by Wright-Fisher sampling according to Eq. (12), or a corresponding master equation. Before presenting the results, we go on a little methodological excursion which will also apply to the later sections on finite populations.

Methods of analysis: For all but very simple selection schemes, it is the hour of the Monte Carlo'ist, especially if dynamical aspects are also considered. Interestingly, although the stochastic equations are more difficult analytically than their deterministic counterparts, they are simpler to handle in simulations. This is because, unless the symmetries of the fitness landscape can be used to advantage (as in, e.g., [164, 142]), an exceedingly large number of configurations must be dealt with in the deterministic case, whereas the number of states to keep track of at any instant is limited by the number of individuals in a finite population, which is usually much smaller.

If generations are discrete and the stochastic component is introduced by Wright-Fisher sampling (12), the corresponding simulations are straightforward. For large populations, the multinomial sampling is quite time consuming, but may be sped up in various ways, e.g. by the sampling scheme in [49] or the multinomial algorithm in [32, p.559]. However, many applications instead use the corresponding continuous-time formulation via a master equation. It lends itself directly to Monte-Carlo simulations. Many authors explicitly reference the paper by Gillespie [59], which gives a very nice and ready-to-use exposition to simulation methods taylored for stochastic chemical reaction systems (note that the population genetic equations may, indeed, be formally understood as reaction systems in a flow reactor; this is actually the context Eigen rediscovered them in [38]).

For simple selection schemes, some analytical approximations of the stochastic equations are possible. They often employ moment expansions [18] for various quantities of interest. To be more precise, both moment [76] and cumulant [145] expansions have proved useful. The notorious problem with both methods is the fact that lower-order

moments or cumulants depend on the respective higher-order quantities, and some kind of closure approximation must be made. Possible choices include the use of higher moments derived from the corresponding deterministic equations [76], as well as maximum entropy considerations. The latter approximation is not very satisfactory, however, if mutation, selection and drift are the only evolutionary forces considered [144]. It works much better when recombination is present, too (recombination is a main ingredient of genetic algorithms, for which the method was originally developed; for review, see [145]). This is because recombination reduces the higher-order cumulants and, with them, the sensitivity of the solution towards them [144].

Results: Error thresholds in finite populations were first described by Nowak and Schuster [135] for the sharply-peaked landscape. Here, the fraction of advantageous sequences is the relevant random variable, and its expectation and variance are of primary interest. In simulations, they may be measured as long-time averages. Interestingly, the expectation follows the deterministic curve for small mutation rates, but then 'jumps' to (near) zero; see Fig. 5. This could also be corroborated analytically [173]. The transition is characterized by large fluctuations. Beyond this point, the population is indistinguishable from a finite population on a flat landscape.

More precisely, finite population size shifts error thresholds to lower mutation rates by an amount which seems to be roughly proportional to $1/\sqrt{N}$ [135]. Apparently, the deterministic solution may be interpreted as the time average of the stochastic process for mutation rates outside the 'window' between the deterministic and stochastic transitions, but this is not so within the window.

Very similar observations hold for the corresponding SPL in the RNA world [148], where one structure is distinguished over all others. For small mutation rates, the favourable phenotype is conserved, while the genotypes diffuse through the corresponding neutral network. At some critical mutation rate, a *phenotypic error threshold* takes place [85, 148].

Simulations were also performed for the Sherrington-Kirkpatrick spin glass [17]. In line with the deterministic results for the closely related Hopfield's landscape, three regimes emerge here. For small mutation rates, the population remains stationary in the vicinity of one high peak; for intermediate mutation rates, it starts wandering across secondary (but more abundant) peaks; and for high mutation rates, it diffuses through all of sequence space in the long run.

5.1.6 Clues from the real world

We have, so far, addressed error thresholds as 'phenomena', thereby implying they are real. However, we have not yet faced the question whether they are relevant problems for evolution. Today's species clearly exist and are genetically well-defined entities. The question must therefore be rephrased to read: Has the mutation rate of these organisms evolved to avoid the error threshold, or didn't it have to bother? It is widely believed that the former is the case, but the experimental evidence is not yet conclusive. Let us follow the clues.

An indirect piece of evidence comes from comparison of genome sizes and mutation rates across species. A roughly inverse relationship is observed for quite a variety of organisms and genome sizes, see the recent survey by Drake at al [37]. This was taken to indicate that the mutation rate has evolved to avoid the error threshold [122] as given by Eq. (27). However, this conclusion implicitly relies on the assumption that the SPL is the relevant landscape for the whole genome; the reservations concerning this assumption have been considered in Sect. 4.2.

More direct evidence comes from mutagenesis experiments with RNA viruses. These viruses have very large genetic variability even at their natural (spontaneous) mutation rate, as reviewed in [34]. If their mutation rate is increased with the help of chemical mutagens, the fitness of the population decreases. The virus does not survive mutation rates larger than twice or three times the spontaneous one, presumably because the error threshold is surpassed; see [80] and [33, 35] for reviews.

Of course, such measurements are rather crude and do not give any hints at the details of the phase transition. More detailed information can only be gained from observation, and possibly sequencing, of large samples, taken from large populations under stationary conditions and a variety of mutation rates. With the power of sequencing methods increasing almost daily, this might not be out of reach for viruses or populations of RNA molecules which may be replicated in the lab under tunable mutation rates; see [16] for a review. If such observations become available, they will tell us a lot about fitness landscapes. After all, we have seen that some fitness landscapes have error thresholds, whereas others do not.

5.2 Muller's ratchet

The prototype model which displays Muller's ratchet is aimed at the multilocus context in *finite* populations. It describes the fate of a finite population which is released at the peak of a multiplicative fitness landscape and experiences deleterious mutations, as well as genetic drift. To be more precise, an infinite number of sites is assumed, each of which may be 'wildtype' (+) or 'mutant' (−), and mutation occurs in a unidirectional fashion (from + to −) at genomic mutation rate U, i.e. the infinite sites limit is assumed. The population is haploid (or diploid without dominance), and the fitness of an individual with j mutations is $w_j = (1-s)^j$. A population of size N undergoes Wright-Fisher sampling (12) at every discrete generation.

Since we have a permutation-invariant landscape, it is sufficient for most considerations to consider classes of individuals with the same number of mutations. Under the

assumptions made, the probability of the j'th mutation class to be sampled reads

$$\psi_j = \frac{1}{N}\sum_{k=0}^{j} v_{j,j-k}\frac{w_{j-k}}{\bar{w}}n_{j-k} \quad \text{where } v_{j,j-k} = \frac{U^k}{k!}e^{-U}\,; \tag{36}$$

cf. Eq. (13). Alternatively, it may be assumed that sampling takes place right after selection (instead of after mutation), in which case one has

$$\psi_j = \frac{1}{N}\sum_{k=0}^{j}\frac{w_j}{\bar{w}}n_j v_{j,j-k}\,. \tag{37}$$

The qualitative behaviour of the model is intuitively clear. Under the action of mutation, some individuals will soon acquire mutant sites, even if the whole population was initially free of mutations. If then, due to the hasards of the sampling process, no individual with zero mutations becomes mother in the next generation, the zero mutation class is lost from the population. This process repeats itself because the then actual fittest class will have the same fate as the zero mutation class. The mean fitness of a clonal lineage will decline incessantly by the successive loss of the actual least loaded classes. Due to the unidirectional nature of mutation, 'better' genotypes can never be reestablished; thus, the mechanism is irreversible and proceeds in a ratchet-like manner.

Recall from Sect. 4.3 that, in the infinite sites limit, novel mutations will always occur at different sites. For large but finite L, this is still the case for most mutations. Accumulation of deleterious mutations may thus occur without necessarily invoking fixation of specific configurations themselves [76, 23]. Here, a fundamental difference between sexual and asexual reproduction becomes apparent. If recombination were present, the fittest genotype could be re-established, except in the rare case that a specific site is fixed for a mutation. Therefore, the ratchet-like deterioration process is much more pronounced with asexual than with sexual reproduction. This was first pointed out by Muller (1964) [131], and Felsenstein (1974) [44] named the process Muller's ratchet.

5.2.1 Ratchet dynamics

It is instructive to consider, for a comparison, the corresponding model of mutation and selection in an *in*finite population, with *finite* L, and *symmetric* mutation according to Eq.(17). We have met this before in the context of sequence space models, and the stationary state (i.e. the distribution of genomes with j mutations) was given by Eq. (28). In the infinite genome limit (24), this converges to a Poisson distribution with parameter U/s (this holds irrespective of whether mutation at single sites is symmetric or unidirectional, provided the population was initially released at the fitness peak). This is termed the *deterministic limit* in the literature on Muller's ratchet, and much of the theory is based upon it, cf. Haigh (1978) [69]. If, instead of (36), (37) is used, one obtains a Poisson distribution with parameter $U(1-s)/s$, instead of U/s [52]. This may be more realistic in certain cases, and we shall adhere to it in what follows.

Let us now move on to finite population size, which will destroy the stationarity of the solution. The crucial events in the process are the losses of the actual least loaded classes – these are known as 'turns' of the ratchet. It is, therefore, important to have good estimates of the rate of the ratchet as a function of the parameters N, s, and U. A crucial quantity is the relative size of the least loaded class. In the deterministic limit, this is

$$p_0 = \exp\Big(- \frac{U(1-s)}{s}\Big) \, . \tag{38}$$

Equation (38) may be used to obtain a very rough estimate for the rate of ratchet: in every generation, it turns with the probability P that all N sampled mothers of the next generation do not belong to the mutation-free class, therefore,

$$P = (1 - p_0)^N . \tag{39}$$

This estimate is accurate only if the process is very slow. For, in this case, the distribution of the mutation classes is restored to the deterministic expectation between any two turns. But even when it is not accurate, it gives some qualitative feel for the ratchet dynamics. Large population size decreases the rate of the ratchet and does so efficiently if $U(1-s)/s$ is not too large. The smaller s or the larger U, the faster turns the ratchet.

If the process is very fast (that is, if $Np_0 < 1$, perhaps because of high mutation rates), the ratchet can be treated as a quasi-deterministic process (for details see [58]).

Many attempts have been made to obtain good estimates of the rate of the ratchet [69, 141, 13, 52, 114, 159, 20, 58, 76, 23], but all approximations are valid only for restricted parameter regimes. The problem is astonishingly difficult for one so simply defined. The main problem is that, under conditions where the ratchet rate is reasonably large, the shape of the expected distribution deviates considerably from the deterministic limit and is hard to predict in a generally valid form. In particular, no simple scaling between U, s, and N seems to exist [76].

5.2.2 Error thresholds versus Muller's ratchet

Error thresholds and Muller's ratchet have a lot in common: They both describe mutational degradation; in the prototype models, this involves delocalization and, in particular, loss of the fittest genotype. Both effects require infinite L. However, error thresholds may be present in infinite populations, whereas Muller's ratchet requires the stochastic component brought about by finite population size. After all, delocalization occurs in the SPL beyond a critical (small) mutation rate, whereas the Poisson distribution which emerges in the deterministic limit of the Muller's ratchet model is localized for *any* $U < 1$. What is the crucial difference?

Wagner and Krall [168] approached the question on the basis of a general model class with infinite population size, discrete generations, unidirectional mutation, and the infinite genome limit (24). Fitness landscapes come from the permutation invariant class, with fitness w_j decreasing monotonically with j, the number of '$-$' sites. The population initially consists of wild type individuals only. In line with the classical localization results ([129, 101]; for review, see [19]) the authors show that the following are equivalent:

a.) The sequence of fitnesses $\{w_j\}$ has no positive lower limit.

b.) For $U < 1$, there is a stationary distribution, with a nonvanishing frequency of the fittest genotype.

Let us remark that a.) requires extensive scaling of fitness.

The result is plausible in that it stresses the purging effect of strongly deleterious mutations. Put differently, a.) entails that a few alleles are sufficiently advantageous to prevent the population from spreading 'too thinly' over the entire genotype space.

The SPL, as the prototype landscape for error thresholds, does not fulfil a.), whereas the multiplicative fitness scheme of the Muller's ratchet model does. As a consequence, the fittest genotype is lost for some $U_c < 1$ with the SPL, but not for the multiplicative scheme, as long as the infinite genome limit is assumed. This explains the need for stochastic effects in the prototype model for Muller's ratchet.

As we have discussed above, however, some care must be exercised in transferring results relying on the infinite genome limit to sequence space models. After all, the latter rely on the *infinite population limit* (25), with its *extensive* scaling of mutation. In order to illustrate this point, let us reconsider the multiplicative fitness function. We have seen that the population remains *localized* in the *infinite genome* limit, as also predicted by a.) and b.). On the other hand, it is *delocalized* in the infinite *population* limit, cf. Eq. (28) and the discussion in section (5.1.3). Thus, the different limits implied, together with the scaling assumed, are also important for differences between error thresholds and Muller's ratchet. However, no general characterization is available at this stage.

5.2.3 Mutational melt-down and extinction

So far, we have assumed that population size remains constant, even throughout the course of fitness deterioration. Constant population size irrespective of the mean fitness is built into the standard sampling process (12), which is invariant under multiplication of all (Wrightian) fitness values with a constant. As long as the number of potentially viable offspring exceeds the carrying capacity of the biotope, this is a reasonable assumption. At some stage of the ratchet process, however, this will no longer hold. Then, *absolute* fitness values will become relevant, and population size will start to decline.

Models of this process were first considered by Lynch and Gabriel (1990) [117]. They require an explicit model of population dynamics in addition to mutation, selection, and drift. In essence, one assumes that the expected number of offspring per individual, R, is independent of fitness and so large that the number of offspring exceeds the carrying capacity C of the biotope at the beginning of every generation. Then, offspring are allowed to survive probability w_j, $0 \leq w_j \leq 1$ (viability fitness). Finally, population regulation (e.g. according to a modified logistic equation) brings the population back to a size less than or equal to C.

Of course, every population will finally die out due to chance effects, even without deleterious mutations, cf. [50, 140]. However, the long time scales involved will not be our concern here. With the assumptions on selection, mutation and drift as in the Muller's ratchet prototype model, mean fitness declines at constant rate until the expected number of viable offspring falls below the carrying capacity (or, equivalently, the expected number of viable offspring per individual, $\langle R \cdot \bar{w} \rangle$, falls below 1). Then the population goes to extinction rapidly because a gradual reduction of population size accelerates the rate of the ratchet. This process is known as *mutational meltdown*. Large fluctuations of population size occur in the vicinity of the extinction point; the extinction times themselves, however, show astonishingly little variation for given sets of parameters [117]. As to parameter dependence, extinction times depend on the time course of fitness deterioration, and this, in turn, depends on the rate of the ratchet as well as on the damage per turn. A change of the detrimental mutational effect s has opposite effects on these. On the one hand, an increase in s implies that selection is more efficient in removing new mutations from the population and the rate of the ratchet is reduced; on the other hand, the damage per turn increases. At small values of s, the second effect predominates because the reduced rate of the ratchet is overcompensated by the fitness loss per turn [117]. An intermediate mutational effect s_{min} minimizes the time to extinction (or maximizes the extinction risk). For $s > s_{\mathrm{min}}$, selection becomes so efficient in removing deleterious mutations that the extinction risk declines [52] (after all, if all mutations are lethal, the ratchet will not turn at all). The dependence of the extinction risk on s is quite strong: At s_{min}, the time to extinction is reduced by several orders of magnitude.

Although we set out to exclude recombination from our considerations, we should like to add that mutational meltdown also endangers small sexual populations (up to a population size of about 1000) [51, 110, 115, 14]. A contrasting result, predicting recombining populations larger than 100 to be safe from Muller's ratchet [25], is attributed [116] to the lack of explicit population dynamics in the simulations.

5.2.4 How to escape Muller's ratchet

In the prototype model, Muller's ratchet, and mutational meltdown as its consequence, are unavoidable. But are they generic features of models of mutation, selection and drift?

In the prototype model, the ratchet may be slowed down by low mutation rate, large population size, and large values of s. But there are less obvious possibilities to retard the ratchet when the assumptions of the prototype model are slightly relaxed. If, instead of equal mutational effects at all sites, a distribution of effects is assumed, longevity is enhanced in a very pronounced way, although the ratchet is not entirely halted [117]. The reason is that, even though the effects of mutation are deleterious on average, the variance

of the mutational effects introduces the possibility that the reduction in fitness caused by a mutation at one locus may be compensated by a beneficial mutation at another locus.

Along another line of thought, positive epistasis was assumed and shown to halt the ratchet entirely [105]. However, this seems to be a nongeneric situation since it is reverted when a distribution of (unconditionally deleterious) effects is additionally introduced [20].

But there are other possibilities to escape the ratchet once one relaxes the assumptions of the prototype model. Instead of the multiplicative fitness function with unidirectional mutation, Wagner and Gabriel [53, 167] considered a model of quantitative genetics with multivariate Gaussian mutation and selection, as introduced in Section 4.2. In this landscape, the proportion of compensatory mutations increases with the distance from the peak. Consequently, the population will, during the process of sliding off the peak, finally arrive at a point where compensatory mutations become relevant, and eventually halt the sliding. A similar result holds for the multiplicative fitness function when mutation is changed from unidirectional to symmetric [76]. Recall that unidirectional mutation is a perfectly reasonable assumption in the infinite sites limit, even if mutation is originally symmetric, provided the population is at the fitness peak. If, however, back mutations are at all present in the original model, they will eventually become important at some distance from the peak.

In both situations, the population reaches a mutation-selection-drift equilibrium, at which it moves in a spherical shell centered at the fittest genotype (or phenotype) [167, 76]. The radius of the shell increases with the mutation rate, but remains stable on a time average. Thus, whereas the decrease of fitness ceases (and Muller's ratchet is halted), the turnover of genotypes continues. If the population is released beneath its equilibrium fitness, it will even climb towards that equilibrium, a process favoured by higher mutation rates – in such cases, a substantial proportion of the occurring mutations is beneficial.

We are thus back at a question about fitness landscapes: How abundant are compensatory mutations?

5.2.5 Clues from the real world

Clonal lineages arise frequently in both the animal and the plant kingdom. Their life spans are fairly short (on the order of $10^4 - 10^5$ generations) and show surprisingly little variation (see [117] and references therein), in line with the predictions of the prototype model. On the other hand, there are a few obligate clonal species which are very old. One example are bdelloid rotifers, tiny multicellular organisms which endure harsh environmental conditions [143]. If Muller's ratchet were the general explanation for the extinction of clonal lineages, why could these species be an exception? One might speculate that, due to the selective pressures acting on them, s is very large. But other 'halting' mechanisms might be considered as well.

The stochastic process of fitness loss over generations was be directly observed in so-called serial passage experiments with viruses, where samples are repeatedly transferred from culture to culture [21]. In these experiments, population size was reduced to the extreme case of one individual between passages. One should be reluctant to speak of Muller's ratchet in this extreme situation – it simply demonstrates that deleterious mutations do occur (Brian Charlesworth, personal communication). Nevertheless, fitness loss is also observed with less severe population bottlenecks, and the dependence on population size has been examined; this is reviewed by Domingo and Holland [35].

5.2.6 Error thresholds and mutational meltdowns

We have, so far, considered explicit population dynamics, and mutational meltdowns as its consequence, in the context of Muller's ratchet only. It is a logical next step to examine them for the prototype model of error thresholds, too. Malarz and Tiggemann [120] introduced explicit population dynamics into a finite population on an SPL, together with sequence space mutation (note that this includes back mutation to the fittest type).

In line with what we know from error thresholds in finite populations (Sect. 5.1.5), there is a stationary distribution of types provided the mutation rate is small. As a consequence, mutational meltdown, too, can only occur when a critical mutation rate is surpassed. Malarz and Tiggemann compared the parameter dependence of this meltdown process with that of the corresponding stochastic error threshold. Perhaps not surprisingly, the meltdown point will only agree with the stochastic error threshold in the unrealistic case in which the expected number of offspring matches the carrying capacity even in a population which has lost all favourable sequences. Otherwise, meltdown takes place at smaller mutation rates [120]. In any case, one meets the typical meltdown curves with accelerated meltdown rate and large fluctuations around the extinction point. The equilibrium structure of the model was corroborated by bifurcation analysis of a toy model in [11], see also Bagnoli and Bezzi, this volume.

6 Connections with the molecular evolutionary process

Much of what we have been concerned with so far was related to equilibrium considerations and/or the time course of (mutational) degeneration. However, some recent interest has turned to the dynamics of adaptation, and to the time course of molecular evolution (i.e. the turnover of genotypes) once a well-adapted state has been reached. Much of the concepts and methods discussed so far also lend themselves to the study of dynamical questions.

One recent study of adaptation dynamics concerns RNA structures. Working with folding algorithms to assign the phenotype and taking selective disadvantage to be proportional to the distance from some target sequence, Huynen et al. [85] demonstrated that evolution proceeds in an intermittent fashion related to the underlying landscape (i.e. neutral networks of frequent structures percolating sequence space and penetrating each other). Long phases of constant phenotype, during which the population diffuses

in the underlying neutral space, alternate with selection-induced, sudden transitions at positions where networks come close to each other. These may be termed *adaptive sweeps* in the language of population genetics, because they temporarily eliminate all molecular variation.

Intermittent behaviour, including hovering about metastable states, seems to be typical of the time evolution on rugged fitness landscapes in general; see, e.g. the analysis of the random energy model by Zhang [179]. In general, if the interval of observation is long enough, a biphasic behaviour seems to be common. A population released at some random point first enters a fast mode with large jumps in fitness; this is hardly affected by population size. Later, it enters a noise-assisted mode with small, rare jumps, for which the finite population size is the driving force. Such or similar observations were reported for the random energy model [179], for the Sherrington-Kirkpatrick spin glass [4], and for the NK model [138]. Here, we meet the old population genetics rule of thumb (cf. Section 3.2): For a new allele with selective advantage s over the fitness of the existing population, selection is in effect if $Ns > 1$; otherwise, random effects predominate.

This 'noise-assisted' phase is the crucial one for considerations of molecular evolution in the long run, with comparisons of different populations or even species in mind. In this context, the loss of correlation between sequences in the course of their divergence from a common ancestor is a quantity of prime interest – after all, this is the observable when sequences from different species are compared. For a strictly flat landscape as originally proposed for the neutral theory, the behaviour is well-known but is at variance with the observations, see [60]. The first not-strictly-neutral models considered were Ohta's *shift models*, assuming an unchanging fraction of deleterious mutations to be available over evolutionary time, with advantageous mutations so rare that they do not contribute significantly to molecular evolution; this is the *nearly neutral theory* as reviewed in [138]. These models owe their name to the fact that incessant fixation of deleterious mutations and, as a consequence, an incessant decline in mean fitness (over phylogenetic time!) is the invariable consequence – a Muller's ratchet like process! For precisely these reasons, the shift models did not stand up to scrutiny and were later replaced by the house-of-cards model (or random energy) model with fitness values drawn from a Gaussian distribution [139, 161]. A detailed analysis was performed by Gillespie [62, 64]. He observed that, under the house-of-cards model, populations evolve towards a state where a large fraction of new mutations are deleterious, but of those which fix (i.e. make it into the population), one-half is advantageous, the other half is deleterious; the reader will recognize this as a variation on the theme 'How can Muller's ratchet be halted?'.

Here, too, the mutations that fix are the many ones which behave nearly neutral, that is, their fitness does not deviate from the mean fitness by more than $1/N$; they are the relevant ones for molecular evolution in the long run. Strongly deleterious mutations hardly ever fix; strongly advantageous ones have become very rare because the current well-adapted state is hard to beat (this may be corroborated by the theory of records).

So far, everything is plausible, and one feels tempted to accept the house-of-cards model as a model of molecular evolution. But there is one startling surprise. Let σ denote the standard deviation of the Gaussian distribution from which the fitness values are drawn. For $N\sigma > 4$, molecular evolution stops [62, 64]! In nature, however, sequence evolution proceeds for population sizes which range over several orders of magnitude, certainly including parameter values with $N\sigma > 4$. Therefore, the house-of-cards model (with fitness values from a Gaussian distribution) cannot be the 'right' model for molecular evolution.

Such a negative result may be a disappointing one to close with. However, one thing is remarkable: it has been achieved to rule out a fitness landscape on the grounds of very general observations as well as advanced theoretical reasoning. This should be encouraging for the theoretician.

7 Further directions

Quite obviously, there are more open problems than solved ones. Let us close by emphasizing one dual pair of questions which we think theoretical physicists could contribute to in the future.

On the theoretical side, it is imperative to *explore more models.* We have seen that the knowledge of the conditions which lead to mutational degradation (through error thresholds, Muller's ratchet, or both) is disappointingly sparse. All examples known so far may be considered as case studies. It would be highly desirable to learn about the behaviour of larger model classes, instead of the singular examples studied so far. Which fitness landscapes exhibit error thresholds, and which display Muller's ratchet? How fast does the ratchet turn? Recall that, even for the prototype model, a full answer is not available. More general answers to these questions could help to decide whether or not mutational degradation should be considered a generic phenomenon.

On the phenomenological side, it is clear that ultimate reasoning about fitness landscapes must be based on real world data. The most powerful data currently available are sequence samples from populations, and from closely related species. It is an important task to *improve methods for inference of evolutionary history from sequence data.* The inference problem we have met in (16) is at the heart of this. It involves simulation of the coalescence process, which are demanding even for neutral evolution. Getting them to work under various assumptions on the fitness landscape is a real challenge. Computational physicists should find a lot to do here.

Acknowledgements

Our views of the subject have been shaped by cooperation and fruitful discussion with a number of physicists, in particular, Michael Baake, Joachim Hermisson, and Holger Wagner, who also helped with constructive criticism on the manuscript. It is our pleasure

to thank Reinhard Bürger, Brian Charlesworth, Joe Felsenstein, Wolfgang Stephan, and Günter Wagner for valuable discussions.

References

[1] Can physics deliver another biological revolution? *Nature* (Opinion), 397:89, 1999.

[2] E. Akin. *The Geometry of Population Genetics*, Springer, New York, 1979.

[3] D. Alves and J. F. Fontanari. A population genetics approach to the quasispecies model. *Phys. Rev. E*, 54:4048, 1996.

[4] C. Amitrano, L. Peliti, and M. Saber. Population dunamics in a spin-glass model of chemical evolution. *J. Mol. Evol.*, 29:513, 1989.

[5] P. W. Anderson. Suggested model for prebiotic evolution: the use of chaos. *Proc. Natl. Acad. Sci. USA*, 80:3386, 1983.

[6] C. F. Aquadro. Insights into the evolutionary process from patterns of DNA sequence variability. *Curr. Op. Gen. Devel.*, 7:835, 1997.

[7] E. Baake. Diploid models on sequence space. *J. Biol. Syst.*, 3:343, 1995.

[8] E. Baake, M. Baake, and H. Wagner. The Ising quantum chain is equivalent to a model of biological evolution. *Phys. Rev. Lett.*, 78:559, 1997; and Erratum, Phys. Rev. Lett. 79 (1997), 1782.

[9] E. Baake, M. Baake, and H. Wagner. Quantum mechanics versus classical probability in biological evolution. *Phys. Rev. E*, 57:1191, 1998.

[10] E. Baake and T. Wiehe. Bifurcations in haploid and diploid sequence space models. *J. Math. Biol.*, 35:321, 1997.

[11] F. Bagnoli and M. Bezzi. Eigen's error threshold and mutational meltdown in a quasispecies model. *Int. J. Mod. Phys. C*, 9:999, 1998.

[12] N. H. Barton and M. Turelli. Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.*, 23:337, 1989.

[13] G. Bell. Recombination and the immortality of the germ line. *J. Evol. Biol.*, 1:67, 1988.

[14] A. T. Bernardes. Mutation load and the extinction of large populations. *Physica A*, 230:156, 1996.

[15] C. K. Biebricher. Replication and evolution of short-chained RNA species replicated by $Q\beta$ replicase. *Cold Spring Harbor Symposia on Quantitative Biology*, LII:299, 1987.

[16] C. K. Biebricher and W. C. Gardiner. Molecular evolution of RNA in vitro. *Biophys. Chem.*, 66:179, 1997.

[17] S. Bonhoeffer and P. F. Stadler. Error thresholds on correlated fitness landscapes. *J. Theor. Biol.*, 164:359, 1993.

[18] R. Bürger. Moments, cumulants, and polygenic dynamics. *J. Math. Biol.*, 30:199, 1991.

[19] R. Bürger. Mathematical properties of mutation-selection models. *Genetica*, 102/103:279, 1998.

[20] D. Butcher. Muller's ratchet, epistasis and mutational effects. *Genetics*, 141:431, 1995.

[21] L. Chao. Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348:454, 1990.

[22] B. Charlesworth. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res. Camb.*, 55:199, 1990.

[23] B. Charlesworth and D. Charlesworth. Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet. Res. Camb.*, 70:63, 1997.

[24] D. Charlesworth, E. E. Lyons, and L. B. Litchfield. Inbreeding depression in two highly inbreeding populations of *leavenworthia*. *Proc. R. Soc. Lond. B*, 258:209, 1994.

[25] D. Charlesworth, M. T. Morgan, and B. Charlesworth. Mutation accumulation in finite outbreeding and inbreeding populations. *Genet. Res. Camb.*, 61:39, 1993.

[26] J. Crow and M. Kimura. The theory of genetic loads. In *Proc. XI Int. Congr. Gen.*, vol. 2, p.495. Pergamon, Oxford, 1964

[27] J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory.* Harper & Row, New York, 1970.

[28] J. F. Crow and M. J. Simmons. The mutation load in *Drosophila.* In M. Ashburner, H. L. Carson, and J. N. Thompson (Eds.), *The genetics and biology of Drosophila*, vol. 3c, p.1. Academic Press, 1983.

[29] B. Derrida. Random energy model: an exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613, 1981.

[30] B. Derrida and D. Bessis. Statistical properties of valleys in the annealed random map model. *J. Phys. A*, 21:L509, 1988.

[31] B. Derrida and L. Peliti. Evolution in a flat fitness landscape. *Bull. Math. Biol.*, 53:355, 1991.

[32] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, 1986.

[33] E. Domingo, C. Escarmis, N. Sevilla, A. Moya, S. F. Elena, J. Quer, I. S. Novella, and J. J. Holland. Basic concepts in RNA virus evolution. *FASEB J.*, 10:859, 1996.

[34] E. Domingo and J. J. Holland. High error rates, population equilibrium, and evolution of RNA replication systems. In Domingo et al. (Eds.), *RNA Genetics*, vol. III:3. CRC Press, Boca Raton, 1988.

[35] E. Domingo and J. J. Holland. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.*, 51:151, 1997.

[36] P. Donnelly and S. Tavaré. Coalescents and the genealogical structure under neutrality. *Annu. Rev. Genet.*, 29:401, 1995.

[37] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow. Rates of spontaneous mutation. *Genetics*, 148:1667, 1998.

[38] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwiss.*, 58:465, 1971.

[39] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasi-species. *Adv. Chem. Phys.*, 75:149, 1989.

[40] J. A. Endler. *Natural selection in the wild*. Princeton University Press, Princeton, 1986.

[41] W. Ewens. *Mathematical Population Genetics*. Springer, New York, 1979.

[42] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.*, 3:87, 1972.

[43] M. Feldman, S. Otto, and F. Christiansen. Population genetics perspectives on the evolution of recombination. *Annu. Rev. Genet.*, 30:261, 1997.

[44] J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78:737, 1974.

[45] J. Felsenstein, M. K. Kuhner, J. Yamato, and P. Beerli. Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. Proceedings of the AMS Summer Workshop on *Statistics in Molecular Biology* (June, 1997), F. Seillier-Moiseiwitsch (Ed.), in press.

[46] S. Franz and L. Peliti. Error threshold in simple landscapes. *J. Phys. A*, 30:4481, 1997.

[47] S. Franz, L. Peliti, and M. Sellitto. An evolutionary version of the random energy model. *J. Phys. A*, 26:L1195, 1993.

[48] Y.-X. Fu and W.-H. Li. Statistical tests of neutrality of mutations. *Genetics*, 133:693, 1993.

[49] W. Gabriel and R. Bürger. Fixation of clonal lineages under Muller's ratchet. submitted.

[50] W. Gabriel and R. Bürger. Survival of small populations under demographic stochasticity. *Theor. Pop. Biol.*, 41:44, 1992.

[51] W. Gabriel, R. Bürger, and M. Lynch. Population extinction by mutational load and demographic stochasticity. In A. Seitz and V. Loeschcke (Eds.), *Species conservation: a population biological approach*, p. 49. Birkhäuser, Basel, 1991.

[52] W. Gabriel, M. Lynch, and R. Bürger. Muller's ratchet and mutational meltdowns. *Evolution*, 47:1744, 1993.

[53] W. Gabriel and G. P. Wagner. Parthenogenetic populations can remain stable in spite of high mutation rate and random drift. *Naturwiss.*, 75:204, 1988.

[54] S. Galluccio. Exact solution of the quasispecies model in a sharply-peaked fitness landscape. *Phys. Rev. E*, 56:4526, 1997.

[55] S. Galluccio, R. Graber, and Y.-C. Zhang. Diffusion on a hypercubic lattice with pinning potential: exact results for the error-catastrophe problem in biological evolution. *J. Phys. A*, 29:L249, 1996.

[56] S. Gavrilets. Evolution and speciation on holey adaptive landscapes. *TREE*, 12:307, 1997.

[57] S. Gavrilets and J. Gravner. Percolation on the fitness hypercube and the evolution of reproductive isolation. *J. Theor. Biol.*, 184:51, 1997.

[58] D. D. G. Gessler. The constraints of finite size in asexual populations and the rate of the ratchet. *Genet. Res. Camb.*, 66:241, 1995.

[59] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403, 1976.

[60] J. Gillespie. *The Causes of Molecular Evolution*. Oxford University Press, New York, 1991.

[61] J. Gillespie. Substitution processes in molecular evolution. II. Exchangeable models from population genetics. *Evolution*, 48:1101, 1994.

[62] J. Gillespie. Substitution processes in molecular evolution. III. Deleterious alleles. *Genetics*, 138:943, 1994.

[63] J. H. Gillespie. Molecular evolution and the neutral allele theory. *Oxford Surveys in Evolutionary Biology*, 9:10, 1987.

[64] J. H. Gillespie. On Ohta's hypothesis: most amino acid substitutions are deleterious. *J. Mol. Evol.*, 40:64, 1995.

[65] R. C. Griffiths and S. Tavaré. Simulating probability distributions in the coalescent. *Theor. Pop. Biol.*, 46:131, 1994.

[66] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, and I. L. Hofacker. Analysis of RNA sequence structure maps by exhaustive enumeration I. Neutral networks. *Monatsh. Chem.*, 127:355, 1996.

[67] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, and I. L. Hofacker. Analysis of RNA sequence structure maps by exhaustive enumeration II. Structures of neutral networks and shape space covering. *Monatsh. Chem.*, 127:375, 1996.

[68] K. P. Hadeler. Stable polymorphisms in a selection model with mutation. *SIAM J. Appl. Math.*, 41:1, 1981.

[69] J. Haigh. The accumulation of deleterious genes in a population - Muller's ratchet. *Theor. Pop. Biol.*, 14:251, 1978.

[70] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part V: Selection and mutation. *Proc. Camb. Phil. Soc.*, 23:838, 1928.

[71] D. L. Hartl and A. G. Clark. *Principles of Population Genetics*. Sinauer, Sunderland, 3rd edition, 1997.

[72] P. G. Higgs. A population genetics model for multiple quantitative traits exhibiting pleiotropy and epistasis. Preprint, 1999.

[73] P. G. Higgs. Error thresholds and stationary mutant distributions in multi–locus diploid genetics models. *Genet. Res. Camb.*, 63:63, 1994.

[74] P. G. Higgs. Frequency distributions in population genetics parallel those in statistical physics. *Phys. Rev. E*, 51:95, 1995.

[75] P. G. Higgs. Compensatory neutral mutations and the evolution of RNA. *Genetica*, 102/103:91, 1998.

[76] P. G. Higgs and G. W. Woodcock. The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. *J. Math. Biol.*, 33:677, 1995.

[77] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125:167, 1994.

[78] J. Hofbauer. The selection mutation equation. *J. Math. Biol.*, 23:41, 1985.

[79] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge, 1988.

[80] J. J. Holland, J. C. de la Torre, and D. A. Steinhauer. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly be chemical mutagenesis. *J. Virol.*, 64:3960, 1990.

[81] D. Houle, D. K. Hoffmaster, S. Assimacopoulos, and B. Charlesworth. The genomic mutation rate for fitness in *drosophila*. *Nature*, 359:58, 1992.

[82] R. R. Hudson. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.*, 7:1, 1990.

[83] R. R. Hudson. How can the low levels of DNA sequence variation in regions of *drosophila* genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA*, 9:6815, 1994.

[84] R. R. Hudson, M. Kreitman, and M. Aguadé. A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116:153, 1987.

[85] M. A. Huynen, P. F. Stadler, and W. Fontana. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397, 1996.

[86] M. O. Johnston and D. J. Schoen. Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science*, 267:226, 1995.

[87] K. S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, London, 1981.

[88] S. Kauffman and S. Johnsen. Coevolution to the edge of chaos: Coupled fitness landscapes, poised states, and coevolutionary avalanches. *J. Theor. Biol.*, 149:467, 1991.

[89] S. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11, 1987.

[90] S. A. Kauffman. Metabolic stability and epigenesis in randomly connected sets. *J. Theor. Biol.*, 22:437, 1969.

[91] S. A. Kauffman. Developmental logic and its evolution. *BioEssays*, 6:82, 1986.

[92] S. A. Kauffman. *The Origin of Order*. Oxford University Press, New York, 1993.

[93] S. A. Kauffman. *At Home in the Universe*. Oxford University Press, New York, 1995.

[94] P. D. Keightley. The nature of deleterious load in *drosophila*. *Genetics*, 144, 1996.

[95] P. D. Keightley and A. Caballero. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *PNAS*, 94:3823, 1997.

[96] M. Kimura. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, 61:893, 1969.

[97] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, 1983.

[98] M. Kimura. *Population Genetics, Molecular Evolution, and the Neutral Theory: Selected Papers*. The University of Chicago Press, Chicago, 1994.

[99] M. Kimura and J. F. Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49:725, 1964.

[100] M. Kimura and T. Maruyama. The mutational load with epistatic gene interactions in fitness. *Genetics*, 54:1337, 1966.

[101] J. F. C. Kingman. On the properties of bilinear models for the balance between genetic mutation and selection. *Math. Proc. Camb. Phil. Soc.*, 81:443, 1977.

[102] J. F. C. Kingman. The coalescent. *Stoch. Proc. Appl.*, 13:235, 1982.

[103] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Prob.*, 19A:27, 1982.

[104] J. Kogut. An introduction to lattice gauge theory and spin systems. *Rev. Mod. Phys.*, 51:656, 1979.

[105] A. S. Kondrashov. Muller's ratchet under epistatic selection. *Genetics*, 136:1469, 1994.

[106] A. S. Kondrashov. Measuring spontaneous deleterious mutation process. *Genetica*, 102/103:183, 1998.

[107] S. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Pop. Biol.*, 51:210, 1997.

[108] M. K. Kuhner, J. Yamato, and J. F. Felsenstein. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics*, 140:1421, 1995.

[109] R. Lande. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Gen. Res. Camb.*, 26:221, 1976.

[110] R. Lande. Mutation and conservation. *Conservation Biology*, 9:782, 1995.

[111] I. Leuthäusser. An exact correspondence between Eigen's evolution model and a two-dimensional Ising system. *J. Chem. Phys.*, 84:1884, 1986.

[112] I. Leuthäusser. Statistical mechanics on Eigen's evolution model. *J. Stat. Phys.*, 48:343, 1987.

[113] W.-H. Li. *Molecular Evolution*. Sinauer, Sunderland, 1997.

[114] M. Lynch, R. Bürger, D. Butcher, and W. Gabriel. The mutational meltdown in asexual populations. *J. Hered.*, 84:339, 1993.

[115] M. Lynch, J. Connery, and R. Bürger. Mutation accumulation and the extinction of small populations. *Amer. Nat.*, 146:489, 1995.

[116] M. Lynch, J. Connery, and R. Bürger. Mutational meltdowns in sexual populations. *Evolution*, 49:1067, 1995.

[117] M. Lynch and W. Gabriel. Mutation load and survival of small populations. *Evolution*, 44:1725, 1990.

[118] M. Lynch and J. B. Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, 1998.

[119] C. A. Macken, P. S. Hagen, and A. S. Perelson. Evolutionary walks on rugged landscapes. *SIAM J. Appl. Math.*, 51:799, 1991.

[120] K. Malarz and D. Tiggemann. Dynamics in Eigen quasispecies model. *Int. J. Mod. Phys. C*, 9:481, 1987.

[121] R. Mañé. *Ergodic theory and differentiable dynamics*. Springer, Heidelberg, 1987.

[122] J. Maynard Smith. *Evolutionary Genetics*. Oxford University Press, Oxford, 1989.

[123] J. Maynard Smith. Estimating the minimum rate of genetic transformation in bacteria. *J. Evol. Biol.*, 7:525, 1994.

[124] J. Maynard Smith, N. H. Smith, M. O'Rourke, and B. G. Spratt. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA*, 90:4384, 1993.

[125] J. Maynard Smith and E. Szathmáry. *The major transitions in evolution*. Freeman, Oxford, 1995.

[126] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.

[127] T. Mitchell-Olds. The molecular basis of quantitative genetic variation in natural populations. *TREE*, 10:324, 1995.

[128] P. A. P. Moran. Random processes in genetics. *Proc. Camb. Phil. Soc.*, 54:60, 1958.

[129] P. A. P. Moran. Global stability of genetic systems governed by mutation and selection. *Math. Proc. Camb. Phil. Soc.*, 80:331, 1976.

[130] T. Mukai, S. I. Chigusa, L. E. Mettler, and J. F. Crow. Mutation rate and dominance genes affecting viability. *Genetics*, 72:335, 1972.

[131] H. J. Muller. The relation of recombination to mutational advance. *Mutat. Res.*, 1:2, 1964.

[132] J. D. Murray. *Mathematical Biology*. Springer, New York, 2nd edition, 1993.

[133] T. Nagylaki. *Introduction to Theoretical Population Genetics*. Springer, New York, 1992.

[134] C. Neuhauser and S. Krone. The genealogy of samples in models with selection. *Genetics*, 145:519, 1997.

[135] M. Nowak and P. Schuster. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J. Theor. Biol.*, 137:375, 1989.

[136] O. Ohnishi. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effects of polygenic mutations. *Genetics*, 87:529–545, 1977.

[137] T. Ohta. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. USA*, 90:10676, 1993.

[138] T. Ohta. Evolution by nearly neutral mutations. *Genetica*, 102/103:83, 1998.

[139] T. Ohta and H. Tachida. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics*, 126:219, 1990.

[140] K. F. Pal. Extinction of small populations in the bit-string model of biological ageing. *Int. J. Mod. Phys. C*, 7:899, 1996.

[141] P. Pamilo, M. Nei, and W. H. Li. Accumulation of mutations in sexual and asexual populations. *Genet. Res.*, 49:135, 1987.

[142] J. R. Peck. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, 137:597, 1994.

[143] G. O. Poinar, Jr., and C. Ricci. Bdelloid rotifers in dominican amber: Evidence for parthenogenetic continuity. *Experientia*, 48:408, 1992.

[144] A. Prügel-Bennett. Modelling evolving populations. *J. Theor. Biol.*, 185:81, 1997.

[145] A. Prügel-Bennett and J. L. Shapiro. The dynamics of a genetic algorithm for simple random Ising systems. *Physica D*, 104:75, 1997.

[146] V. A. Ratner, A. A. Zharkikh, N. Kolchanov, S. N. Rodin, V. V. Solovyov, and A. S. Antonov. *Molecular Evolution*. Springer, Berlin, 1995.

[147] M. Reed and B. Simon. *Functional Analysis*, volume I. Academic Press, San Diego, 1980.

[148] C. Reidys, C. Forst, and P. Schuster. Replication and mutation on neutral networks. Bull. Math. Biol., in press.

[149] C. Reidys, P. Stadler, and P. Schuster. Generic properties of combinatory maps: Neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 59:339, 1997.

[150] C. M. Reidys and P. F. Stadler. Neutrality in fitness landscapes. Preprint, 1998.

[151] W. R. Rice. Degeneration of a nonrecombining chromosome. *Science*, 1994.

[152] D. Ruelle. *Statistical Mechanics: Rigorous Results*. Addison-Wesley, Redwood City, 1983.

[153] D. S. Rumschitzky. Spectral properties of Eigen evolution matrices. *J. Math. Biol.*, 24:667, 1987.

[154] P. Schuster and J. Swetina. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.*, 50:635, 1988.

[155] P. Stadler and R. Happel. Random field models for fitness landscapes. J. Math. Biol., in press.

[156] P. F. Stadler. Landscapes and their correlation functions. *J. Math. Chem.*, 20:1, 1996.

[157] D. A. Steinhauer and J. J. Holland. Rapid evolution of RNA viruses. *Annu. Rev. Microbiol.*, 41:409, 1987.

[158] W. Stephan. The rate of compensatory evolution. *Genetics*, 144:419, 1996.

[159] W. Stephan, L. Chao, and J. G. Smale. The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genet. Res. Camb.*, 61:225, 1993.

[160] D. L. Swofford, G. J. Olsen, P. J. Waddell, and D. M. Hillis. Phylogenetic inference. In D. M. Hillis, C. Moritz, and B. K. Mable (Eds.), *Molecular Systematics*, p. 407. Sinauer, Sunderland, 1995.

[161] H. Tachida. A study on a nearly neutral mutation model in finite populations. *Genetics*, 128:183, 1991.

[162] M. Tacker, P. F. Stadler, E. G. Bornberg-Bauer, I. L. Hofacker, and P. Schuster. Algorithm independent properties of RNA secondary structure predictions. *Eur. Biophys. J.*, 25:115, 1996.

[163] F. Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123:585, 1989.

[164] P. Tarazona. Error threshold for molecular quasispecies as phase transition: From simple landscapes to spin glass models. *Phys. Rev.*, A45:6038, 1992.

[165] C. J. Thompson. *Mathematical Statistical Mechanics.* Macmillan, New York, 1972.

[166] C. J. Thompson and J. L. McBride. On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math. Biosci.*, 21:127, 1974.

[167] G. P. Wagner and W. Gabriel. Quantitative variation in finite populations: what stops Muller's ratchet in the absence of recombination? *Evolution*, 44:715, 1990.

[168] G. P. Wagner and P. Krall. What is the difference between models of error thresholds and Muller's ratchet. *J. Math. Biol.*, 32:33, 1993.

[169] H. Wagner, E. Baake, and T. Gerisch. Ising quantum chain and sequence evolution. *J. Stat. Phys.*, 92:1017, 1998.

[170] G. Weiss. The impact of evolutionary models on the inference of population history. In M. K. Uyenoyama and A. von Haeseler, editors, *Proceedings of the Trinational Workshop on Molecular Evolution*, pages 113–122, Durham, 1998. Duke Univ. Publ.

[171] M. C. Whitlock, P. C. Philips, F. B.-G. Moore, and S. J. Tonsor. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.*, 26:601, 1995.

[172] T. Wiehe. Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.*, 69:127, 1997.

[173] T. Wiehe, E. Baake, and P. Schuster. Error Propagation in Reproduction of Diploid Organisms. *J. Theor. Biol.*, 177:1, 1995.

[174] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proc. Sixth Int. Congr. Genet.*, volume 1, pages 356–366, 1932.

[175] S. Wright. *Evolution and the genetics of populations*, volume 1. Genetic and biometric foundations. University of Chicago Press, Chicago, 1968.

[176] S. Wright. *Evolution and the genetics of populations*, volume 2. The theory of gene frequencies. University of Chicago Press, Chicago, 1969.

[177] S. Wright. *Evolution and the genetics of populations*, volume 3. Experimental results and evolutionary deductions. University of Chicago Press, Chicago, 1977.

[178] S. Wright. *Evolution and the genetics of populations*, volume 4. Variability within and among natural populations. University of Chicago Press, Chicago, 1978.

[179] Y.-C. Zhang. Quasispecies evolution of finite populations. *Phys. Rev. E*, 55:R3817, 1997.

[180] M. Zuker and D. Sankoff. RNA secondary structures and their prediciton. *Bull. Math. Biol.*, 46:591–621, 1984.

[181] R. Zurmühl. *Praktische Mathematik für Ingenieure und Physiker.* Springer, 5 th edition, 1965. Reprint 1984.

Figure captions

Figure 1:
The sampling process and the coalescent process. In a population of constant size, individuals are sampled with replacement to contribute offspring to the next generation according to Eq. (12). This may be viewed as a bifurcation process forward in time, or a coalescent process backward in time. The genealogy of a sample of $M = 4$ alleles is indicated (fat lines).

Figure 2:
The genealogy of the sample in Fig. 1. Usually, the history of the sample is unknown. Then, the genealogy is a random variable, with the T_m denoting the lengths of the time intervals during which there are m distinct lineages in the genealogy.

Figure 3:
Stationary state of the paramuse model (32) and (33) for sequences of length $L = 30$, and the SPL with selective advantage $s = 0.03$ of the favourable sequence. Relative frequencies p_j of sequences with j '$-$' sites are represented as a height profile over the xy plane with j ($0 \leq j \leq 30$) in the x direction, and the mutation rate per site, μ ($0 \leq \mu \leq 0.002$) in the y direction. For mutation rates above the error threshold ($\mu_c \simeq 0.001$), the population is evenly distributed over sequence space, i.e. $p_j = \frac{1}{2^L}\binom{L}{j}$.

Figure 4:
Sequence genealogy as a two-dimensional, anisotropic Ising model. Sequences descending from each other (i.e. grandmother-mother-daughter) form the rows of the lattice. This way, columns correspond to sequence positions, rows to generations, and the transition from one row to the next is governed by the mutation-reproduction matrix $\mathcal{VW}$. If the letters of the sequence are identified with spins, every genealogy corresponds to one possible configuration of a two-dimensional Ising model, and the transfer matrix takes the role of the mutation-reproduction matrix. Interactions are anisotropic: Interactions *between* the rows (dashed) are nearest neighbour and correspond to mutation. Interactions *within* the rows (solid lines) may be arbitrary and long-range, but are identical for every row. Their presence or absence, as well as their strengths, define the fitness landscape in the sense of a mapping from sequence space into the real numbers.

Figure 5:
Error thresholds in finite populations. Long-term average of the relative frequency of the favourable sequence, $\langle N_1/N \rangle$, as a function of the mutation rate μ. Fitness landscape: SPL with selective advantage $s = 0.05$ of the fittest sequence; and sequence length: $L = 30$. Solid line: $N = \infty$ (deterministic limiting case); dashed: $N = 10000$; dotted: $N = 1000$; dash-dotted: $N = 100$.

Figure 6:

Muller's ratchet and mutational meltdown (schematically). Fat line: Time course of the average number of surviving offspring per individual, $\langle R \cdot \bar{w} \rangle$ (R is the maximum number of offspring and $\bar{w}$ the mean viability). $\langle R \cdot \bar{w} \rangle$ declines due to Muller's ratchet; when it falls below 1, the population goes to extinction rapidly (mutational meltdown). Thin line: histogram of extinction probabilities. Extinction times show surprisingly small coefficients of variation.

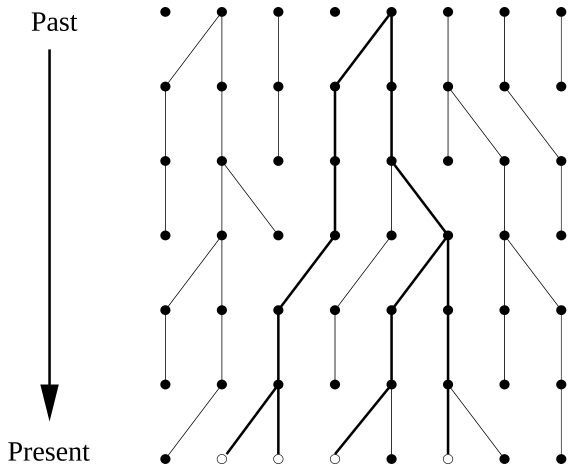

Figure 1:

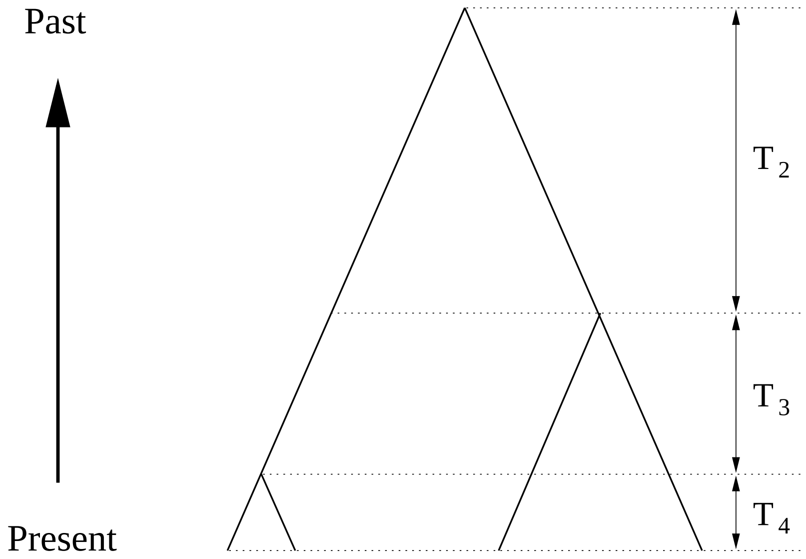

Figure 2:

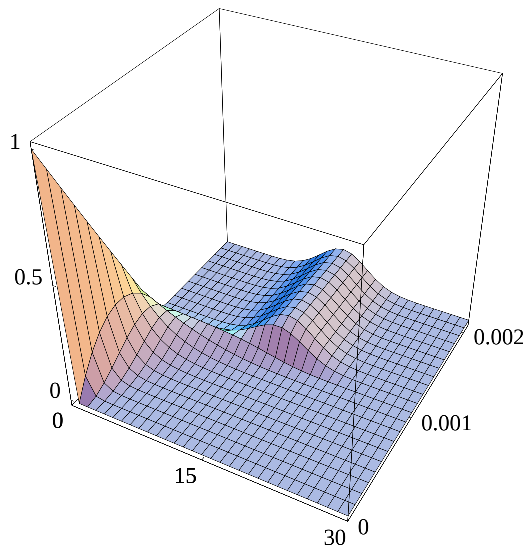

Figure 3:

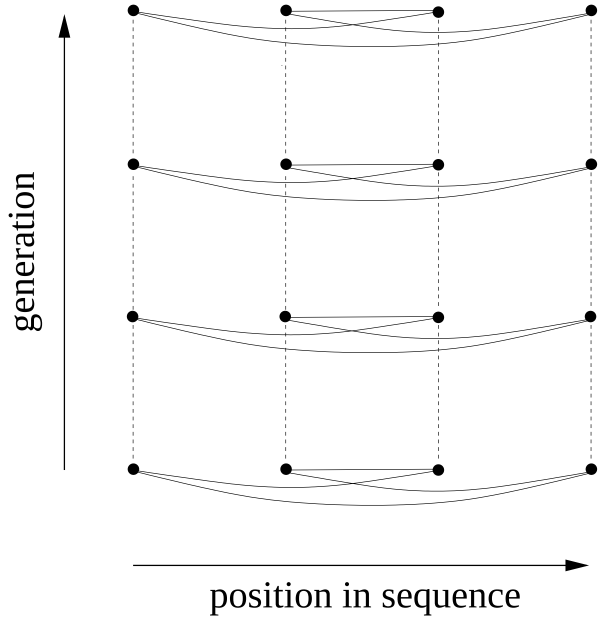

Figure 4:

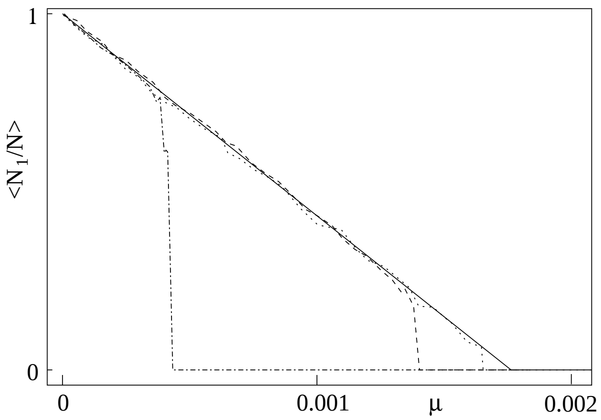

Figure 5:

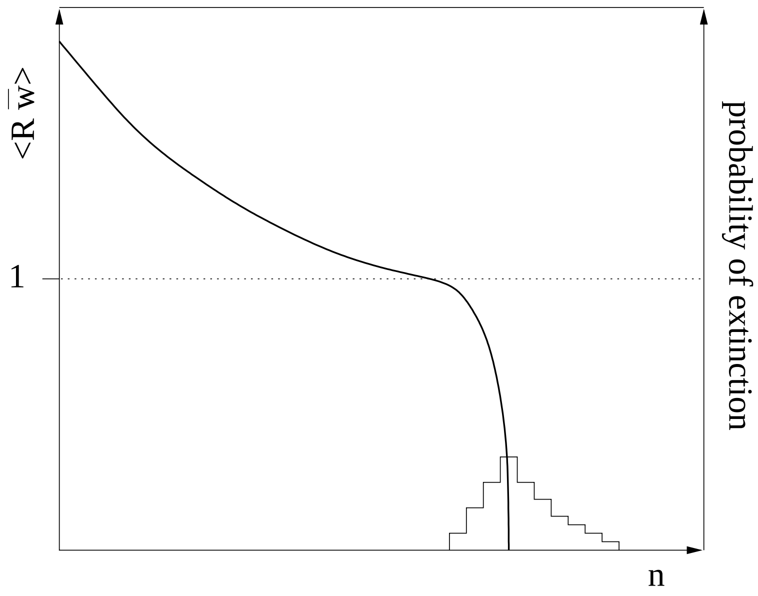

Figure 6: