

Chapter 1

Rate of Adaptation of Large Populations

Feng Yu and Alison Etheridge

Abstract We consider the accumulation of beneficial and deleterious mutations in asexual populations. The rate of adaptation is affected by the total mutation rate, proportion of beneficial mutations, and population size N . We describe two models: a strong selection model for large population size and a weak selection model for moderate population size. For the strong selection model, we give an argument to show that regardless of mutation rates, as long as the proportion of beneficial mutations is strictly positive, the adaptation rate is at least $\mathcal{O}(\log^{1-\delta} N)$, if the population size is sufficiently large. For the weak selection model, we use a Girsanov transform to perform some preliminary calculations that shed new light on the adaptation rate.

1.1 Background and Introduction

How does natural selection affect the rate of adaptation (defined to be the average speed at which mean fitness of the population increases)? To be more specific, let us assume we have a finite population of individuals, labelled $i = 1, \dots, N$, each accumulating μ mutations per generation on average (each mutation is assumed to occur at a different locus). A mutation is beneficial with probability q and deleterious with probability $1 - q$. In this work, we model all mutations to have the same absolute “fitness effects” σ , but their sign may be positive or negative depending on whether the mutation is beneficial or deleterious. We denote the selection coefficient σ , instead of s as in common in the biology literature, because we want to reserve s for a time variable that we use for calculations later on. Refer

F. Yu
School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK
feng.yu@bristol.ac.uk

A. Etheridge
Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK
etheridg@stats.ox.ac.uk

to the description of model A in Sect. 1.2.1 and model B in Sect. 1.2.2 for the exact meaning of fitness effects $\pm\sigma$. A more sophisticated model should consider mutations that have a distribution of fitness effects, e.g. an independent exponentially distributed selective advantage associated with each new beneficial mutation as proposed by Gillespie (1991). Recent numerical work by Hegreness et al. (2006), however, suggests that in models where beneficial mutations have a distribution of fitness advantages, dynamics of adaptation can be reasonably described by an equivalent model where all beneficial mutations confer the same fitness advantage.

Without the effects of selection, we expect the rate of adaptation to be $\mu(2q-1)$. If $q < 1/2$, then the rate of adaptation is in fact negative, meaning that the population is becoming less fit with time. Selection should of course increase the rate of adaptation, but by how much? By Fisher's fundamental theorem (Fisher 1930), the adaptation rate is roughly proportional to the variance of the fitness among individuals in the population. It is well known that selection reduces the variance of fitness, but by how much and through what mechanisms this reduction occurs is not well understood mathematically. This understanding is a necessary step towards understanding the advantage of sex and recombination. Specifically, selection causes linkage disequilibrium and thus reduces the variance of fitness, whereas recombination reduces the linkage disequilibrium and thus increases the variance. The two mechanisms combine to induce a "recombination/selection balance".

Understanding the mechanisms through which selection and mutation interact with each other to determine the variance of fitness among individuals in the population and its adaptation rate also contributes to answering the following separate but related question: Does there exist an upper limit to the rate of adaptation in asexual populations? This question was originally posed to us by Nick Barton and can be more easily described in the context of purely beneficial mutations, i.e. take $q = 1$ in the last paragraph. Let us first consider the simplest scenario where a single beneficial mutation arises in an otherwise neutral population and no further mutations occur until the fate of that mutant is known. This situation is well understood, both in terms of p_{fix} , the fixation probability of the mutation, and the length of time it takes this process to complete. The value of p_{fix} was settled by Haldane (1927), who showed that if the selection coefficient associated with the mutation is σ , then under these circumstances $p_{fix} \approx 2\sigma$. In this case p_{fix} is almost independent of the population size, N .

If the mutation does fix, the process whereby it increases in frequency from $1/N$ to 1 is known as a *selective sweep*. The duration of a selective sweep is $\mathcal{O}(\log(\sigma N)/\sigma)$ generations. In particular, the length of a selective sweep is long if the population size N is large. If one assumes that the mutation rate per individual per generation is μ , then the overall mutation rate will be proportional to population size and we see that for large populations the assumption that no new mutation will arise during the time course of the sweep breaks down. Instead one expects multiple overlapping sweeps. Because there is no recombination, of all the mutations present in the population, only those on one lucky individual who eventually becomes the common ancestor of all individuals in the population will fix. Gerrish and Lenski (1998) called this effect *clonal interference*: clones that carry different beneficial

mutations compete with each other and interfere with one another's growth in the population.

If the process has run long enough and is in some sort of stationarity, we can assume p_{fix} to be the same for all mutations, all of which are by assumption beneficial with fitness effect $+\sigma$. Furthermore, all mutations must eventually either fix or become extinct in finite time. Then the rate of adaptation is proportional to $\mu N p_{fix}$, where μN is the total number of mutations that occur to all individuals in the population in a single generation. If p_{fix} is independent of population size, then we expect an adaptation rate of $\mathcal{O}(N)$. But as the population size N becomes larger, the total number of mutations extant in the population at any particular time also grows, and thus the clonal interference will be more severe, reducing p_{fix} . This leads to the following question: If one does not limit the number of simultaneous selective sweeps, what is p_{fix} , or equivalently, what is the rate of adaptation? As $N \rightarrow \infty$, is the rate of adaptation finite or does it increase without bound? Barton and Coe (2007) suggest that there is an asymptotic limit to the rate of adaptation. Other authors (e.g. Rouzine et al. 2003; Wilke 2004; Desai and Fisher 2007) argue that no such limit exists.

Previous work on this question has adopted two general approaches: (1) calculate the fixation probability p_{fix} directly and (2) study the distribution of fitness of all individuals in the population and ask how this distribution evolves with time. The first approach was used in Gerrish and Lenski (1998), Wilke (2004) and Barton and Coe (2007), amongst others. This approach does not seem to be easy to adapt into a rigorous mathematical argument. Instead, we follow the second approach, i.e. to consider the distribution of fitnesses in the population, as in Rouzine et al. (2003, 2007) and Desai and Fisher (2007).

We will follow Rouzine et al. (2003) in taking fitness effects to be additive and all selection coefficients to be equal; thus, an individual's fitness can be characterised by the *net* number of beneficial mutations which it carries (which may be negative). Writing P_k for the proportion of individuals with fitness equivalent to k beneficial mutations, $\{P_k\}_{k \in \mathbb{Z}}$ forms a type of *travelling wave* whose shape remains basically unchanged over time. The position of the wave moves to the left or the right on the fitness axis, depending on whether the adaptation rate is positive or negative. The shape of the wave fluctuates stochastically even after a long time, and the fluctuations are larger if the population size is smaller. So the approximation of this *stochastic* wave by a *fixed* shape is most accurate when the population size is large. The approach is to treat the "bulk" of the wave as deterministic but the leading edge (i.e. the size of the fittest class) as stochastic. They found that the rate of adaptation, which is equal to the speed of the travelling wave, asymptotically depends logarithmically on population size N , which is consistent with results of *in vitro* studies of a type of RNA virus in Novella et al. (1995, 1999). Rouzine et al. (2007) offers an improved treatment of the stochastic edge.

Desai and Fisher (2007) followed a similar but slightly different approach. Just as in Rouzine et al. (2003), they studied the adaptation rate by studying how the sizes of different fitness classes change with time, and they also divided classes into deterministic and stochastic regimes. But instead of studying the wave speed directly, they examined the variance of fitness and studied how much variance can be maintained by mutation while the population is acted on by selection.

We have so far concentrated on beneficial mutations. In reality, most mutations are either neutral or deleterious. In particular, if all mutations in an asexual population were deleterious, the population would irreversibly accumulate deleterious mutations, a process known as Muller’s ratchet. The first rigorous analysis of this phenomenon was due to Haigh (1978). Despite a very considerable body of work on Muller’s ratchet, e.g. Higgs and Woodcock (1995), Stephan et al. (1993), Gordo and Charlesworth (2000) and Etheridge et al. (2007), a rigorous expression for the rate of decline in mean fitness of the population remains elusive. In this work, we are interested in both this rate of decline in mean fitness and the question of whether a large population can overcome Muller’s ratchet.

The conclusion we reach in Sect. 1.3 is the following: as long as the proportion of beneficial mutations is strictly positive, the rate of adaptation is roughly $\mathcal{O}(\log N)$ for large N , where N is the population size. Cuthbertson et al. (2007) established a rigorous lower bound of any fractional power of $\log N$ for the adaptation rate. This shows that no matter how small the proportion of beneficial mutations is, a large enough population size will yield a positive adaptation rate and Muller’s ratchet is overcome. It also shows, in particular, that the rate of adaptation grows without bound as $N \rightarrow \infty$ in the all-mutations-beneficial case. This is consistent with the findings of Rouzine et al. (2003), Wilke (2004) and Desai and Fisher (2007). Furthermore, Barton and Coe (2007) argue that the adaptation rate in the case of unlinked loci and thus no linkage disequilibrium (i.e. links between loci are disrupted by very large recombination rates) is $\mathcal{O}(\log N)$. If this is true, then recombination can only increase the adaptation rate within the same order.

Throughout our work, we use continuous-time Moran models, where each individual has an exponentially distributed lifetime and mutations fall on each individual at a rate of μ . In the biological literature one would expect to see a Wright–Fisher model, where each individual has a lifetime of exactly one generation and fitness is often taken to be *multiplicative*. In the Moran model, individuals accumulate mutations during their lifetime, rather than at reproduction as in the Wright–Fisher model, but the average number of mutations accumulated by an individual during one unit of time in the Moran model can be made to correspond to the average number of mutations accumulated by an individual at each reproduction event in the Wright–Fisher model. Fitness in the context of Moran models is taken to be *additive* instead of multiplicative. For large populations, we expect that the results we obtain for our (much more mathematically tractable) Moran model will mirror those for the corresponding Wright–Fisher model.

Figure 1.1 shows a plot of the adaptation rate against \log population size from simulation results of model A described in Sect. 1.2.1. We observe that for each set of parameters q , μ and σ , the rate of adaptation is roughly proportional to $\log N$ and small population sizes may result in negative adaptation rates. Furthermore, larger q results in a higher adaptation rate for fixed μ and σ . The upshot is that with μ and σ held constant, the smaller the proportion of beneficial mutations, the larger the population size required for Muller’s ratchet to be overcome.

The argument we present in Sect. 1.3, as well as the argument used in Rouzine et al. (2003), works well only for large N , where the shape of the travelling wave can

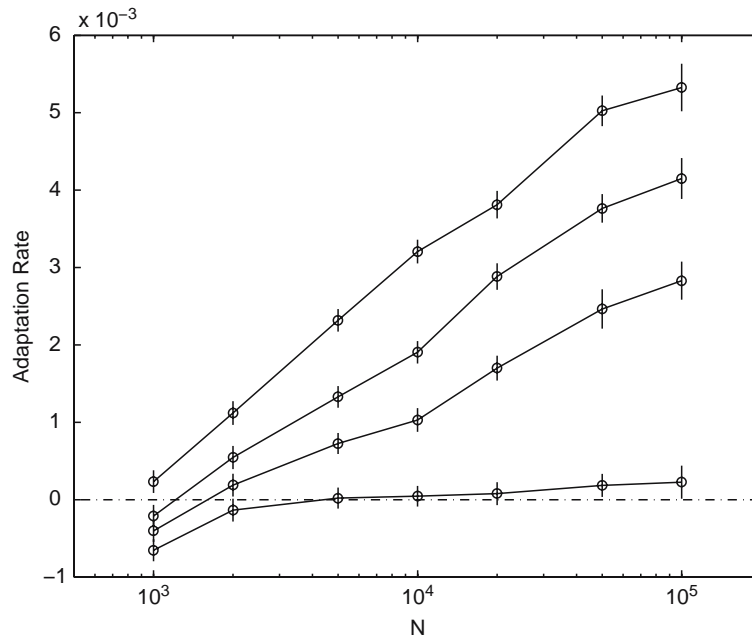


Fig. 1.1 Adaptation rate against population size, from top to bottom, for $q = 4, 2, 1$ and 0.2% , $\mu = 0.01$ and $\sigma = 0.01$. Circles represent actual data points and vertical bars represent 95% confidence intervals

be treated as deterministic. For moderate N , its shape actually fluctuates and an approximation by a fixed shape becomes unsatisfactory. See Fig. 1.2 for a comparison of the travelling wave for population sizes of $N = 1,000$ and $N = 100,000$; the shape for $N = 100,000$ is far more stable than that of $N = 1,000$. In the case of moderate population sizes, it may be more useful mathematically to rescale the mutation and selection parameters with population size to obtain a process where the proportions P_k of “individuals” in fitness class k can take any value in the interval $[0, 1]$. Even though the limiting process is an infinite population limit, its behaviour still resembles a finite population process since by holding μN and σN constant and rescaling time by N , we maintain genetic drift as we take the limit $N \rightarrow \infty$. A limit we obtain this way is often called a *weak selection* model. This is what we do in Sect. 1.2.2. Because the sizes of essentially all fitness types are stochastic, studying the effects of mutation and selection on the adaptation rate requires a different approach and little mathematical analysis has so far been carried out. In Sect. 1.4, we show how to obtain the weak selection model from a neutral one with a Girsanov transform and then perform some calculations that shed new light on the adaptation rate. The aim of working with a weak selection model is still to understand the effect of mutation and selection on the adaptation rate, but this time for moderate population sizes.

The work is organised as follows. In Sect. 1.2, we formulate two models, a strong selection individual-based model and a weak selection diffusion model, and perform

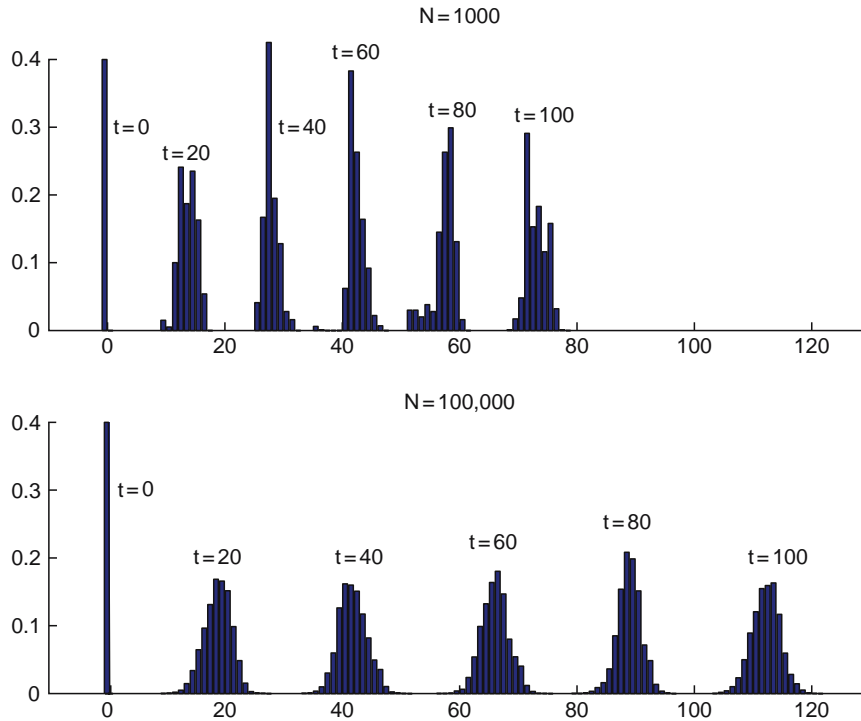


Fig. 1.2 The travelling wave for two populations sizes: $N = 1,000$ and $N = 100,000$. Both plots have parameters $q = 1$, $\mu = 0.02$ and $\sigma = 0.02$

some preliminary calculations on both models. In Sect. 1.3, we present a non-rigorous argument that leads to an asymptotic adaptation rate of roughly $\mathcal{O}(\log N)$ for the strong selection model. And finally in Sect. 1.4, we introduce the Girsanov transform to obtain the weak selection model, and show how it can yield useful information on the adaptation rate. Throughout this work, we focus on mathematical ideas that have the potential of being turned into rigorous arguments, rather than rigorous arguments themselves, which appear elsewhere.

1.2 Two Models

We first describe a finite population Moran model: the strong selection model. Then in Sect. 1.2.2 we describe a weak selection model in which mutation and selection coefficients are taken to be $\mathcal{O}(1/N)$. As $N \rightarrow \infty$, if time is measured in units of size N , the model converges to an infinite-dimensional diffusion.

We assume constant population size N . Let $X_i(t) \in \mathbb{Z}$, $i = 1, \dots, N$, denote the *fitness type* of the i th individual, defined to be the number of beneficial mutations

minus the number of deleterious mutations carried by the individual. We assume all mutations have the same absolute fitness effects; thus, the exact fitness of an individual is determined by its fitness type. Let $P_k(t)$ denote the proportion of individuals that have fitness k at time t , i.e.

$$P_k = \frac{1}{N} \sum_{i=1}^N \delta_{X_i=k}. \quad (1.1)$$

For $p \in \mathcal{P}^N(\mathbb{Z})$, the space of probability measures p on \mathbb{Z} formed by N point masses each with weight $1/N$, define $p_k = p(\{k\})$ and

$$P_{[k,l]} = \sum_{i=k}^l p_i, \quad m_n(p) = \langle k^n, p \rangle = \sum_{k \in \mathbb{Z}} k^n p_k,$$

$$m(p) = m_1(p), \quad c_n(p) = \sum_{k \in \mathbb{Z}} (k - m(p))^n p_k.$$

In particular, $m(p)$ is the mean fitness of the population, and $c_2(p) = m_2(p) - m(p)^2$ is the second central moment of the population fitness, i.e. its variance.

The model of interest is one where each individual accumulates beneficial mutations at a Poisson rate $q\mu$ and deleterious mutations at rate $(1-q)\mu$. We assume a so-called infinitely-many-sites model where each mutation is assumed to be new and occur at a different locus on the genome.

1.2.1 Strong Selection Model

In the strong selection model, all individuals experience mutation, selection (which introduces a drift reflecting the differential reproductive success based on fitness) and genetic drift via resampling. This model, which we call *model A*, is described below:

1. *Mutation*: For each individual i a mutation event occurs at rate μ . With probability $1-q$, X_i changes to $X_i - 1$ and with probability q , X_i changes to $X_i + 1$.
2. *Selection*: For each pair of individuals (i, j) , at rate $\frac{\sigma}{N}(X_i - X_j)^+$, individual i replaces individual j .
3. *Resampling*: For each pair of individuals (i, j) , at rate $\frac{1}{2N}$, individual i replaces individual j .

This model has a time scale such that one unit of time corresponds roughly to one generation. Often one combines the resampling and selection into a single term. Each pair of individuals is involved in a reproduction event at some constant rate and the effect of selection is then that it is more likely to be the fitter individual that reproduces. In fact, this is the selection mechanism in model B, which we describe in Sect. 1.2.2. Since σ is typically rather small (one expects $N\sigma$ to be $\mathcal{O}(1)$) our formulation in model A is a very small perturbation of the model where resampling and selection are combined into a single mechanism, and does not affect our results.

The population $X_i(t)$ forms the empirical measure $P_k(t)$ by Eq. 1.1. We can also describe the mechanisms of model A in terms of the P_k 's:

1. *Mutation*: For any $k \in \mathbb{Z}$, at rate $(1-q)\mu NP_k$, P_k decreases by $\frac{1}{N}$ and P_{k-1} increases by $\frac{1}{N}$; at rate $q\mu NP_k$, P_k decreases by $\frac{1}{N}$ and P_{k+1} increases by $\frac{1}{N}$.
2. *Selection*: For any pair of $k, l \in \mathbb{Z}$ such that $k > l$, at rate $\sigma(k-l)NP_k P_l$, P_k increases by $\frac{1}{N}$ and P_l decreases by $\frac{1}{N}$.
3. *Resampling*: For any pair of $k, l \in \mathbb{Z}$, at rate $\frac{N}{2}P_k P_l$, P_k increases by $\frac{1}{N}$ and P_l decreases by $\frac{1}{N}$.

We perform a so-called martingale decomposition for P_k in Eq. 1.2 below, i.e. we write P_k as a sum of deterministic drifts and martingales, where, roughly, the drift describes the average evolution of P_k and the martingales describe how much the evolution can differ from the average behaviour. For details of this type of calculation, see Ikeda and Watanabe (1981). For this work, we just point out that a pure jump Markov process that increases by $1/N$ at rate $r(t)$ decomposes into a drift term $\frac{1}{N} \int_0^t r(s) ds$ and a martingale term that has quadratic variation $\frac{1}{N^2} \int_0^t r(s) ds$.

$$\begin{aligned}
P_k(t) &= P_k(0) + \mu \int_0^t [q(P_{k-1}(s) - P_k(s)) + (1-q)(P_{k+1}(s) - P_k(s))] ds \\
&\quad + \sigma \int_0^t \sum_{l \in \mathbb{Z}} (k-l) P_k(s) P_l(s) ds + M_k^{P,1}(t) + M_k^{P,2}(t) \\
&= P_k(0) + \mu \int_0^t (qP_{k-1}(s) - P_k(s) + (1-q)P_{k+1}(s)) ds \\
&\quad + \sigma \int_0^t \sum_{l \in \mathbb{Z}} (k-l) P_k(s) P_l(s) ds + M_k^{P,1}(t) + M_k^{P,2}(t), \tag{1.2}
\end{aligned}$$

where $M_k^{P,1}$ and $M_k^{P,2}$ are orthogonal martingales, the first arising from the (compensated) mutation mechanism and the second from the resampling and (compensated) selection mechanisms. Both have maximum jump size $1/N$ and their quadratic variations are all $\mathcal{O}(1/N)$, i.e. $[M_k^{P,1}](t) = \mathcal{O}(1/N)$ and $[M_k^{P,2}](t) = \mathcal{O}(1/N)$. For exact formulae for these quadratic variations, see Cuthbertson et al. (2007).

We can write the martingale decomposition of the mean $m(P(t)) = \sum_k k P_k(t)$ as follows:

$$\begin{aligned}
m(P(t)) &= m(P(0)) + M^{P,m}(t) + \int_0^t \sigma \sum_{k,l \in \mathbb{Z}} k(k-l) P_k(s) P_l(s) \\
&\quad + \mu \sum_{k \in \mathbb{Z}} k (qP_{k-1}(s) - P_k(s) + (1-q)P_{k+1}(s)) ds \tag{1.3} \\
&= m(P(0)) + M^{P,m}(t) + \int_0^t \sigma [m_2(P(s)) - m(P(s))^2] \\
&\quad + \mu \sum_{k \in \mathbb{Z}} [q(k+1) - 1 + (1-q)(k-1)] P_k(s) ds \\
&= m(P(0)) + M^{P,m}(t) + \int_0^t \sigma [c_2(P(s)) + \mu(2q-1)] ds,
\end{aligned}$$

where $M^{P,m}$ is a martingale whose quadratic variation is $\mathcal{O}(1/N)$. In differential notations, we write

$$dm(P) = [\mu(2q - 1) + \sigma c_2(P)] dt + dM^{P,m}. \quad (1.4)$$

1.2.2 Weak Selection Model

For the weak selection model, we let $\tilde{\mu}$ and $\tilde{\sigma}$ denote the mutation and selection coefficients, respectively, and define

$$\tilde{\mu} = N\mu, \quad \tilde{\sigma} = N\sigma,$$

where μ and σ are the mutation and selection coefficients in the strong selection model defined in Sect. 1.2.1. To obtain the mechanisms of the weak selection model, we hold $\tilde{\mu}$ and $\tilde{\sigma}$ constant (i.e. scale μ and σ down by N), combine the selection and resampling mechanism into a single reproduction mechanism, and then speed up time by N , to obtain the following Moran particle model with constant population size N , which we call *model B*:

1. *Mutation*: For each individual i a mutation event occurs at rate $\tilde{\mu}$. With probability $1 - q$, X_i changes to $X_i - 1$ and with probability q , X_i changes to $X_i + 1$.
2. *Reproduction*: For a pair of individuals (i, j) , a reproduction event occurs at rate 1. With probability $\frac{1}{2}(1 + \frac{\tilde{\sigma}}{N}(X_i - X_j))$, individual i replaces individual j ; with probability $\frac{1}{2}(1 - \frac{\tilde{\sigma}}{N}(X_i - X_j))$, individual j replaces individual i .

In one unit of time, each individual experiences roughly N reproduction events; thus, one unit of time corresponds to N generations in this model, and consequently $\tilde{\mu}$ and $\tilde{\sigma}$ correspond with $N\mu$ and $N\sigma$, respectively. Of course, the definition of this reproduction mechanism runs into trouble if any two individuals i and j have fitness difference larger than $N/\tilde{\sigma}$, i.e. if $|X_i - X_j| > N/\tilde{\sigma}$, in which case one of the two probabilities in the reproduction mechanism becomes negative and the other becomes larger than 1. The probability of this happening will converge to 0 as $N \rightarrow \infty$ and we ignore this possibility in the calculations below.

We use $P_k^{(N)}$ to denote the proportion of individuals of fitness type k , to emphasise the fact that this is an empirical measure formed by N individuals each with mass $1/N$, and that the process is a jump process with jump size $1/N$. We will eventually take an infinite population limit $N \rightarrow \infty$ and obtain a process with continuous paths. In terms of the proportions $P_k^{(N)}$'s, model B can be described as follows:

1. *Mutation*: For any $k \in \mathbb{Z}$, at rate $(1 - q)\tilde{\mu}NP_k^{(N)}$, $P_k^{(N)}$ decreases by $\frac{1}{N}$ and $P_{k-1}^{(N)}$ increases by $\frac{1}{N}$; at rate $q\tilde{\mu}NP_k^{(N)}$, $P_k^{(N)}$ decreases by $\frac{1}{N}$ and $P_{k+1}^{(N)}$ increases by $\frac{1}{N}$.
2. *Reproduction*: For any pair of $k, l \in \mathbb{Z}$, at rate $\frac{N^2}{2}(1 + \frac{\tilde{\sigma}}{N}(k - l))P_k^{(N)}P_l^{(N)}$, $P_k^{(N)}$ increases by $\frac{1}{N}$ and $P_l^{(N)}$ decreases by $\frac{1}{N}$; at rate $\frac{N^2}{2}(1 - \frac{\tilde{\sigma}}{N}(k - l))P_k^{(N)}P_l^{(N)}$, $P_l^{(N)}$ increases by $\frac{1}{N}$ and $P_k^{(N)}$ decreases by $\frac{1}{N}$.

Just as for model A, we can perform the martingale decomposition to obtain

$$\begin{aligned}
P_k^{(N)}(t) &= P_k^{(N)}(0) + \tilde{\mu} \int_0^t \left(qP_{k-1}^{(N)}(s) - P_k^{(N)}(s) + (1-q)P_{k+1}^{(N)}(s) \right) ds \\
&\quad + \tilde{\sigma} \int_0^t \sum_{l \in \mathbb{Z}} (k-l) P_k^{(N)}(s) P_l^{(N)}(s) ds + M_k^{(N),P,1}(t) + M_k^{(N),P,2}(t), \\
&= P_k^{(N)}(0) + \tilde{\mu} \int_0^t \left(qP_{k-1}^{(N)}(s) - P_k^{(N)}(s) + (1-q)P_{k+1}^{(N)}(s) \right) ds \\
&\quad + \tilde{\sigma} \int_0^t \left(k - m(P^{(N)}(s)) \right) P_k^{(N)}(s) ds + M_k^{(N),P,1}(t) + M_k^{(N),P,2}(t),
\end{aligned}$$

where $M_k^{(N),P,1}$ and $M_k^{(N),P,2}$ are orthogonal martingales, the first arising from the (compensated) mutation mechanism and the second from the reproduction mechanism. Because we have sped up time by a factor of N , the quadratic variation of $M^{(N),P,2}$ (which we recall arises from the resampling) is no longer of $\mathcal{O}(1/N)$:

$$\begin{aligned}
[M_k^{(N),P,1}](t) &= \mathcal{O}(1/N), \\
[M_k^{(N),P,1}, M_l^{(N),P,1}](t) &= 0 \text{ if } |k-l| \geq 2, \\
[M_k^{(N),P,2}](t) &= \int_0^t \sum_{l \in \mathbb{Z}} P_k^{(N)}(s) P_l^{(N)}(s) ds \\
&= \int_0^t P_k^{(N)}(s) (1 - P_k^{(N)}(s)) ds, \\
[M_k^{(N),P,2}, M_l^{(N),P,2}](t) &= - \int_0^t P_k^{(N)}(s) P_l^{(N)}(s) ds \text{ if } k \neq l.
\end{aligned}$$

Because $M^{(N),P,2}$ is $\mathcal{O}(1)$, it persists even after taking the limit $N \rightarrow \infty$, while $M^{(N),P,1}$ goes away since it is $\mathcal{O}(1/N)$.

We take the limit $N \rightarrow \infty$ to obtain a process $P_k(t)$ that is continuous for each $k \in \mathbb{Z}$ and $t \geq 0$ and satisfies the martingale problem

$$\begin{aligned}
P_k(t) &= P_k(0) + \tilde{\mu} \int_0^t \left(qP_{k-1}(s) - P_k(s) + (1-q)P_{k+1}(s) \right) ds \\
&\quad + \tilde{\sigma} \int_0^t \left(k - m(P(s)) \right) P_k(s) ds + M_k^P(t),
\end{aligned} \tag{1.5}$$

where M_k^P are martingales with quadratic variation process

$$[M_k^P, M_{k'}^P](t) = \int_0^t P_k(s) (\delta_{kk'} - P_{k'}(s)) ds, \tag{1.6}$$

where $\delta_{kk'} = 1$ if $k = k'$ and $\delta_{kk'} = 0$ otherwise.

The existence and uniqueness of the solution of this martingale problem is non-trivial, but it will appear elsewhere and we do not go into detail here. The martingales M_k^P can be represented in terms of Brownian motions in the following way:

$$M_k^P(t) = \int_0^t \sum_{l \in \mathbb{Z}} \sqrt{P_k(s)P_l(s)} dW_{kl}(s), \quad (1.7)$$

where $\{W_{kl} : k, l \in \mathbb{Z}, k > l\}$ are independent Brownian motions and we define $W_{kl} = -W_{lk}$ for $k, l \in \mathbb{Z}$ and $k < l$ and $W_{kk} \equiv 0$ for $k \in \mathbb{Z}$. We now verify that M_k^P has the quadratic variation process specified by Eq. 1.6. We write $P_k = P_k(s)$ if this causes no confusion. In differential notation,

$$d[M_k^P, M_{k'}^P] = \sum_{l, l'} \sqrt{P_k P_l P_{k'} P_{l'}} d[W_{kl}, W_{k'l'}]$$

for $k \neq l$. Since $k \neq l$, $d[W_{kl}, W_{k'l'}] = -dt$ if $k = l'$ and $l = k'$ and $d[W_{kl}, W_{k'l'}] = 0$ otherwise. Therefore,

$$d[M_k^P, M_{k'}^P] = -\sqrt{P_k P_l P_l P_k} dt = -P_k P_l dt.$$

Similarly,

$$\begin{aligned} d[M_k^P] &= \sum_{l, l'} \sqrt{P_k^2 P_{k'} P_{l'}} d[W_{kl}, W_{k'l'}] \\ &= \sum_l \sqrt{P_k^2 P_l^2} d[W_{kl}, W_{kl}] + \sum_{l, l': l \neq l'} \sqrt{P_k^2 P_l P_{l'}} d[W_{kl}, W_{k'l'}]. \end{aligned}$$

Since $d[W_{kl}, W_{kl}] = dt$ if $l \neq k$, and if $l \neq l'$ then $d[W_{kl}, W_{k'l'}] = 0$, we have

$$d[M_k^P] = \sum_{l \neq k} P_k P_l dt = P_k(1 - P_k) dt.$$

Thus, we can write the solution to the martingale problem (Eq. 1.5) as the solution to the following infinite system of stochastic differential equations (SDEs),

$$dP_k = [\tilde{\mu}(qP_{k-1} - P_k + (1-q)P_{k+1}) + \tilde{\sigma}(k - m(P))P_k] dt + \sum_{l \in \mathbb{Z}} \sqrt{P_k P_l} dW_{kl}, \quad (1.8)$$

and then associate an infinite-dimensional operator $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$ with this system:

$$\begin{aligned} \mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}} f(p) &= \sum_k [\tilde{\mu}(qp_{k-1} - p_k + (1-q)p_{k+1}) + \tilde{\sigma}(k - m(p))p_k] \frac{\partial f}{\partial p_k} \\ &\quad + \frac{1}{2} \sum_{k, l} p_k (\delta_{kl} - p_l) \frac{\partial^2 f}{\partial p_k \partial p_l}. \end{aligned} \quad (1.9)$$

The operator $\mathcal{A}_{\bar{\sigma}, q, \bar{\mu}} f(p)$ is usually called the generator of the stochastic process and its domain can be taken to be $C^2(\mathcal{P}(\mathbb{Z}))$, the space of twice continuously differentiable functions on $\mathcal{P}(\mathbb{Z})$, the space of probability measures on \mathbb{Z} . For convenience, we will refer to the stochastic process that solves Eq. 1.8 as the process associated with the generator $\mathcal{A}_{\bar{\sigma}, q, \bar{\mu}}$.

If we set $\bar{\sigma} = 0$, then the process associated with $\mathcal{A}_{0, q, \bar{\mu}}$ solves the SDE

$$dP_k = [\bar{\mu}(qP_{k-1} - P_k + (1-q)P_{k+1})] dt + \sum_{l \in \mathbb{Z}} \sqrt{P_k P_l} dW_{kl},$$

and can be thought of as the *neutral* process that has only the mutation (both beneficial and deleterious) and neutral resampling mechanisms. Because the resampling mechanism is neutral, many techniques, e.g. the look-down construction (Donnelly and Kurtz 1999), can be brought to bear on this process and consequently many quantities can be calculated for the neutral process. The main technique we use to study the selected process $\mathcal{A}_{\bar{\sigma}, q, \bar{\mu}}$ is to write it as a Girsanov transform of the neutral process $\mathcal{A}_{0, q, \bar{\mu}}$. We will discuss this technique in more detail in Sect. 1.4.

1.3 Strong Selection: Asymptotic Adaptation Rate

In this section, we take large population size and focus on the strong selection model A that we described in Sect. 1.2.1. We give a non-rigorous argument that leads to an asymptotic adaptation rate of roughly $\mathcal{O}(\log N)$, as long as q is strictly positive and regardless of the selection and mutation parameters.

The basic idea of this approach is to treat the ‘‘bulk’’ of the travelling wave (where $P_k = \mathcal{O}(1)$) as deterministic and only treat the front and tail of this wave (where $P_k \ll 1$) as stochastic. This is similar to the approach taken in Rouzine et al. (2003), where an approximate expression for the shape of wave is first derived using a deterministic equation. There is an infinite family of solutions to this deterministic equation, parameterised by the wave speed. To determine the correct wave speed for a given population size, stochasticity at the front of the wave must be taken into account. In our work, we derive a set of moment equations and show that the selection mechanism dictates that the form of the wave is approximately Gaussian. We then calculate the speed of this wave in two ways, one using the bulk, the other using the front. These two speeds must be the same, which gives us a constraint that determines our approximation to the adaptation rate.

With the martingale term $M^{P,1}$ and $M^{P,2}$ of $\mathcal{O}(1/\sqrt{N})$, the effect of noise on P_k can be considered to be quite small if P is much larger than $1/N$. For P in this range, we have from Eq. 1.2,

$$\begin{aligned} dP_k &\approx \left[\mu(qP_{k-1} - P_k + (1-q)P_{k+1}) + \sigma \sum_{l \in \mathbb{Z}} (k-l)P_k P_l \right] dt \\ &= [\mu(qP_{k-1} - P_k + (1-q)P_{k+1}) + \sigma(k-m(P))P_k] dt. \end{aligned} \quad (1.10)$$

This is similar to Eq. 2 in Rouzine et al. (2003), except that they treated discrete time. With P_k evolving according to Eq. 1.10, we can write the evolution of the mean fitness $m(P)$ in the following way:

$$dm(P) = (\mu(2q - 1) + \sigma c_2(P)) dt, \quad (1.11)$$

which is simply Eq. 1.4 on dropping the martingale term $M^{P,m}$. We can use the above two equations to calculate the evolution of $c_n(P) = \sum_k (k - m(P))^n p_k$:

$$\begin{aligned} \frac{dc_n(P)}{dt} &= \sum_k (k - m(P))^n [\mu(qP_{k-1} - P_k + (1 - q)P_{k+1}) + \sigma(k - m(P))P_k] \\ &\quad - \sum_k n(k - m(P))^{n-1} (\mu(2q - 1) + \sigma c_2(P))P_k \\ &= \mu \sum_k (k - m(P))^n [qP_{k-1} - P_k + (1 - q)P_{k+1}] - n(k - m(P))^{n-1} (2q - 1)P_k \\ &\quad + \sigma(c_{n+1} - nc_{n-1}c_2) \\ &= \mu \sum_k (k - m(P))^{n-1} [q(k + 1 - m(P)) - (k - m(P))] \\ &\quad + (1 - q)(k - 1 - m(P)) - n(2q - 1)]P_k + \sigma(c_{n+1} - nc_{n-1}c_2). \end{aligned}$$

Thus,

$$\frac{dc_n(P)}{dt} = \sigma(c_{n+1}(P) - nc_{n-1}(P)c_2(P)). \quad (1.12)$$

Notice that Eq. 1.12 does not depend on the mutation coefficient μ , and thus the mutation mechanism does not affect the overall shape of the wave. Since the martingale term $M^{P,m}$ that we dropped in Eq. 1.4 to obtain Eq. 1.11 is small compared with P_k where P_k is significantly larger than $1/N$, we can assume the shape of the wave, as described by the central moments c_n , evolves approximately deterministically according to Eq. 1.12. Setting the right-hand side of Eq. 1.12 to zero for all $n \geq 2$, we see that the shape of P is approximately Gaussian, although Eq. 1.12 admits any Gaussian distribution as its solution regardless of its variance. In other words, Eq. 1.12 has an infinite family of solutions, each a Gaussian distribution, and the system itself yields no information on the exact value of the variance.

It is the behaviour at the front and tail of the wave that determines the value of this variance (and hence the wave speed). Suppose then that P is approximately Gaussian with mean fitness $m(P)$ and variance b^2 . The population size N is large but finite so the ‘‘front’’ of the wave is approximately where the level of P reaches $1/N$ (i.e. one individual). This level is roughly at $K + m(P)$ where K satisfies

$$\frac{1}{2\pi b^2} e^{-K^2/2b^2} = \frac{1}{N},$$

which implies that

$$K \approx b\sqrt{2\log N}. \quad (1.13)$$

We assume that a single individual is born at site $K + m(P)$ at time $t = 0$ and try to find out how long it takes for an individual to be born at site $K + 1 + m(P)$. This is how long it takes the wave to advance by one. Let $Z(t)$ be the number of individuals of the fittest type, that is the size of the population at site $K + m(P)$, then Eq. 1.10 says that until a beneficial mutation falls on site $K + m(P)$, $Z(t)$ increases exponentially at rate $\sigma K - \mu$, ignoring beneficial mutations occurring to type $K - 1 + m(P)$, i.e.

$$Z(t) \approx e^{(\sigma K - \mu)t}. \quad (1.14)$$

While the number of individuals at site $K + m(P)$ grows, all of them are accumulating beneficial and deleterious mutations at rates $q\mu$ and $(1 - q)\mu$, respectively. As soon as one of these individuals accumulates a beneficial mutation, the ‘‘front’’ of the wave will have advanced by one. The probability that no beneficial mutation occurs to any individuals with fitness $K + m(P)$ is

$$\exp\left(-q\mu \int_0^t Z(s) ds\right) = \exp\left(-\frac{q\mu}{\sigma K - \mu} \left(e^{(\sigma K - \mu)t} - 1\right)\right),$$

using the expression for $Z(t)$ in Eq. 1.14. Therefore, the average time when the front advances by one can be calculated to be

$$\frac{e^{q\mu/(\sigma K - \mu)}}{\sigma K - \mu} \int_{1/(\sigma K - \mu)}^{\infty} \frac{e^{-q\mu u}}{u} du.$$

Using the fact that for small ε , $\int_{\varepsilon}^{\infty} e^{-q\mu u}/u du = \log(1/\varepsilon) + \mathcal{O}(1)$, we deduce that the average time until the front advances by one is roughly $\frac{1}{\sigma K - \mu} \log(\sigma K - \mu)$ for large K , which gives a wave speed of $(\sigma K - \mu)/\log(\sigma K - \mu)$.

But by Eq. 1.11, the wave speed is $\mu(2q - 1) + \sigma c_2(P) = \mu(2q - 1) + \sigma b^2 \approx \mu(2q - 1) + \sigma K^2/(2\log N)$, using Eq. 1.13. This leads to the following consistency condition:

$$\frac{\sigma K - \mu}{\log(\sigma K - \mu)} = \mu(2q - 1) + \frac{\sigma K^2}{2\log N}.$$

For large K , this approximately reduces to

$$K \log(\sigma K) = 2\log N.$$

This is a transcendental equation with no closed-form solution. If $K = \log N$, then the left-hand side is greater than the right-hand side; if, on the other hand, $K = \log^{1-\delta} N$ for any small positive δ , then left-hand side is less than the right-hand side. So K is between $\log N$ and any fractional power of $\log N$, i.e. a bit smaller than $\mathcal{O}(\log N)$, which implies a wave speed, that is a rate of adaptation, of order between $\log N$ and any fractional power of $\log N$.

In Cuthbertson et al. (2007), we in fact established a rigorous asymptotic lower bound of $\log^{1-\delta} N$ (where δ is any small positive number) for the adaptation rate, as long as q is strictly positive and regardless of the selection and mutation parameters. The proof uses some of the ideas outlined in this section.

1.4 Weak Selection: Girsanov Calculations

Now we turn to the weak selection model of Sect. 1.2.2. We would like to study the adaptation rate of the *selected* process associated with generator $\mathcal{A}_{\bar{\sigma}, q, \bar{\mu}}$ (defined in Eq. 1.9) with $\bar{\sigma} \neq 0$. We can get a set of moment equations similar to Eq. 1.12, which is not closed and is therefore difficult to study. The *neutral* process associated with $\mathcal{A}_{0, q, \bar{\mu}}$, on the other hand, is much easier to study. The central idea of our approach is to find a relationship between the probability measure associated with the selected process and that associated with the neutral process. The technique we use is the Girsanov transform.

1.4.1 The Girsanov Transform

The Girsanov theorem tells us how the stochastic process changes if we change the underlying probability measure. We review this briefly for the simplest case of adding drift to a one-dimensional standard Brownian motion. Suppose $W(t)$ is a Brownian motion under the probability measure \mathbb{P} . Let $\{\mathcal{F}_t\}_{t \geq 0}$ be the corresponding filtration. Let $a(t)$ be an \mathcal{F}_t -adapted process, then $Z(t) = \exp\{\int_0^t a(s) dB(s) - \frac{1}{2} \int_0^t a(s)^2 ds\}$ is a local martingale. If $a(t)$ satisfies the Novikov condition, i.e. $E[\exp\{\int_0^t a(s)^2 ds\}] < \infty$, then $Z(t)$ is actually a martingale and we can define a probability \mathbb{Q} by

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = Z(t).$$

By the Girsanov theorem (see Karatzas and Shreve 1991), the process

$$X(t) = X(0) - \int_0^t a(s) ds + B(t)$$

is a Brownian motion under \mathbb{Q} . This gives us a way of relating the distribution of the standard Brownian motion and that of a Brownian motion with drift.

The Girsanov theorem works equally well when the Brownian motion described above is replaced with more general continuous martingales or the dimension is larger than 1 but finite. For the infinite-dimensional setting, we can use Dawson's Girsanov theorem (see Dawson 1991 and Theorem 7.9 of Etheridge 2000) developed for Dawson–Watanabe superprocesses. We first recall the *neutral* model that we obtained in Sect. 1.2.2 as the infinite population limit of processes described by

model B. Setting $\tilde{\sigma} = 0$ in Eq. 1.8, the neutral process satisfies the following infinite system of SDEs:

$$dP_k = [\tilde{\mu}(qP_{k-1} - P_k + (1-q)P_{k+1})] dt + dM_k, \quad (1.15)$$

where we define the martingale

$$M_k(t) = \int_0^t \sum_{l \in \mathbb{Z}} \sqrt{P_k(s)P_l(s)} dW_{kl}(s).$$

We associate this process with its generator $\mathcal{A}_{0,q,\tilde{\mu}}$, with $\mathcal{A}_{\tilde{\sigma},q,\tilde{\mu}}$ defined in Eq. 1.9 for any $\tilde{\sigma}$. Let \mathbb{P}_0 denote the law of $\{P_k, k \in \mathbb{Z}\}$ evolving according to the neutral generator $\mathcal{A}_{0,q,\tilde{\mu}}$, then by definition, M_k is a martingale under \mathbb{P}_0 . We define the function $a_n(p)$

$$a_n(p) = \tilde{\sigma} \left(n - \sum_k k p_k \right) = \tilde{\sigma}(n - m(p))$$

and Z to be a process that satisfies the following SDE:

$$dZ = Z \sum_n a_n dM_n = \tilde{\sigma} Z \sum_n (n - m(p)) dM_n. \quad (1.16)$$

In particular, if the Novikov condition is satisfied, i.e. if

$$\mathbb{E}^{\mathbb{P}_0} \left[\exp \left(\tilde{\sigma}^2 \int_0^T c_2(P(s)) ds \right) \right] < \infty,$$

then Z is a martingale up to time T under \mathbb{P}_0 . The Novikov condition does hold for the process $\mathcal{A}_{0,q,\tilde{\mu}}$ under the stationary measure of the process centred about its mean, but we skip the details, which will appear elsewhere. We define

$$\begin{aligned} \tilde{M}_k(t) &= M_k(t) - \tilde{\sigma} \int_0^t \sum_n a_n(s) d[M_k, M_n](s) \\ &= M_k(t) - \tilde{\sigma} \int_0^t \sum_n (n - m(P(s))) (\delta_{kn} - P_k(s)) P_n(s) ds \\ &= M_k(t) - \tilde{\sigma} \int_0^t \left[(k - m(P(s))) P_k(s) - \sum_n (n - m(P(s))) P_k(s) P_n(s) \right] ds \\ &= M_k(t) - \tilde{\sigma} \int_0^t (k - m(P(s))) P_k(s) ds, \end{aligned} \quad (1.17)$$

where we use Eq. 1.6 in the second line. By Dawson's Girsanov theorem, \tilde{M}_k is a $\mathbb{P}_{\tilde{\sigma}}$ -martingale, where

$$\left. \frac{d\mathbb{P}_{\tilde{\sigma}}}{d\mathbb{P}_0} \right|_{\mathcal{F}_t} = Z(t)$$

and

$$[M_k, M_l](t) = [\tilde{M}_k, \tilde{M}_l](t).$$

Although they are martingales under different probability measures, the \tilde{M}_k 's and M_k 's have the same cross-variation structures; thus as in Eq. 1.7 for M_k , we can define $\{\tilde{W}_{kl} : k, l \in \mathbb{Z}, k > l\}$ to be independent Brownian motions, $\tilde{W}_{kl} = -\tilde{W}_{lk}$ for $k, l \in \mathbb{Z}$ and $k < l$, and $\tilde{W}_{kk} \equiv 0$ for $k \in \mathbb{Z}$. Then

$$\tilde{M}_k^P(t) = \int_0^t \sum_{l \in \mathbb{Z}} \sqrt{P_k(s)P_l(s)} d\tilde{W}_{kl}(s) \quad (1.18)$$

are martingales under \mathbb{P}_σ . From the definition of \tilde{M} in Eq. 1.17, we see that M satisfies

$$dM_k = d\tilde{M}_k(t) + \tilde{\sigma}(k - m(P(t)))P_k(t) dt. \quad (1.19)$$

From Eq. 1.15, we easily obtain $dM_k = dP_k - [\tilde{\mu}(qP_{k-1} - P_k + (1-q)P_{k+1})] dt$. We plug this into the left-hand side of Eq. 1.19 and Eq. 1.18 into its right-hand side to obtain

$$dP_k = [\tilde{\mu}(qP_{k-1} - P_k + (1-q)P_{k+1}) + \tilde{\sigma}(k - m(P))P_k] dt + \sum_{l \in \mathbb{Z}} \sqrt{P_k P_l} d\tilde{W}_{kl}.$$

We see this is exactly the SDE in Eq. 1.8 with generator $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$, except the Brownian motions W_{kl} are replaced with another family of Brownian motions \tilde{W}_{kl} .

To summarise, W_{kl} 's are Brownian motions under \mathbb{P}_0 , which is the law for the process P associated with $\mathcal{A}_{0, q, \tilde{\mu}}$, while \tilde{W}_{kl} 's are Brownian motions under \mathbb{P}_σ , the law for P associated with $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$. The Radon–Nikodym derivative $(d\mathbb{P}_\sigma/d\mathbb{P}_0)|_{\mathcal{F}_t}$ is given by $Z(t)$, which is a martingale that satisfies Eq. 1.16 and can be written out explicitly:

$$Z(t) = \exp \left(\int_0^t \sum_k a_k(P(s)) dM_k(s) - \frac{1}{2} \int_0^t \sum_{k, l} a_k(P(s)) a_l(P(s)) d[M_k, M_l](s) \right).$$

We now find a more convenient expression for $Z(t)$. The first integral in the exponent is

$$\begin{aligned} \int_0^t \sum_k a_k(P(s)) dM_k(s) &= \tilde{\sigma} \int_0^t \sum_{k, l} (k - m(P(s))) \sqrt{P_k(s)P_l(s)} dW_{kl}(s) \\ &= \tilde{\sigma} \int_0^t \sum_k k \sum_l \sqrt{P_k(s)P_l(s)} dW_{kl}(s) = \tilde{\sigma} \sum_k k M_k(s), \end{aligned}$$

where the second equality comes from the fact $\sum_{k, l} \sqrt{P_k(s)P_l(s)} dW_{kl}(s) = 0$, since $W_{kl} = -W_{lk}$. By Eq. 1.15,

$$\begin{aligned} \sum_k k M_k(t) &= \sum_k k P_k(t) - \sum_k k P_k(0) - \int_0^t \sum_k k [\tilde{\mu}(qP_{k-1}(s) - P_k(s) + (1-q)P_{k+1}(s))] ds \\ &= m(P(t)) - m(P(0)) - \tilde{\mu}(2q-1)t. \end{aligned}$$

The second integral in the exponent is

$$\begin{aligned}
& \int_0^t \sum_{k,l} a_k(P(s)) a_l(P(s)) d[M_k, M_l](s) \\
&= \tilde{\sigma}^2 \int_0^t \sum_{k,l} (k - m(P(s)))(l - m(P(s))) (\delta_{kl} - P_k(s)) P_l(s) ds \\
&= \tilde{\sigma}^2 \int_0^t \sum_k (k - m(P(s)))^2 P_k(s) ds = \tilde{\sigma}^2 \int_0^t c_2(P(s)) ds.
\end{aligned}$$

Therefore,

$$Z(t) = \exp\left(\tilde{\sigma}[m(P(t)) - m(P(0)) - \tilde{\mu}(2q-1)t] - \frac{\tilde{\sigma}^2}{2} \int_0^t c_2(P(s)) ds\right),$$

where P solves the neutral SDE (Eq. 1.15).

1.4.2 Moments of the Neutral Process

For the calculations on the selected process (Eq. 1.8) that we perform in Sect. 1.4.3, we need to know some asymptotic moments of the neutral process (Eq. 1.15). We drop the notational dependence on p of the moments, m , c_2 , c_3 , etc. The mean of P obeys

$$\begin{aligned}
dm &= \tilde{\mu} \sum_k k[(qP_{k-1} - P_k + (1-q)P_{k+1})] dt + dM \\
&= \tilde{\mu}(2q-1) dt + dM,
\end{aligned} \tag{1.20}$$

where we define the martingale

$$M(t) = \sum_k kM_k(t),$$

which has quadratic variation

$$\begin{aligned}
d[M] &= \sum_{k,l,k',l'} k \sqrt{P_k P_l} k' \sqrt{P_{k'} P_{l'}} d[W_{kl}, W_{k'l'}] \\
&= \sum_{k,l,k',l':k'=k,l'=l} kk' \sqrt{P_k P_l P_{k'} P_{l'}} dt - \sum_{k,l,k',l':k=l',k'=l} kk' \sqrt{P_k P_l P_{k'} P_{l'}} dt \\
&= \sum_{k,l} k^2 P_k P_l dt - \sum_{k,l,k',l'} kl P_k P_l dt \\
&= (m_2 - m^2) dt \\
&= c_2 dt.
\end{aligned}$$

Shiga (1982) showed that although there does not exist a stationary distribution for the neutral process $P(t)$, there does exist a stationary distribution for the process

centred about its mean, and furthermore, the centred process is ergodic, i.e. after a long time, the process behaves like the stationary distribution. Therefore, we expect the expectation of all the central moments of P to converge to something as $t \rightarrow \infty$. To find the exact value of $\lim_{t \rightarrow 0} E^{\mathbb{P}^0}[c_n(P(t))]$, $n = 2, 3, \dots$, we use the generator $\mathcal{A}_{0,q,\tilde{\mu}}$ to calculate its effect on c_n . First, it can be verified that for $n \geq 2$,

$$\begin{aligned}\frac{\partial c_n}{\partial p_k} &= (k-m)^n - knc_{n-1} \\ \frac{\partial^2 c_n}{\partial p_k \partial p_l} &= -ln(k-m)^{n-1} - nk((l-m)^{n-1} - (n-1)lc_{n-2}),\end{aligned}$$

where $c_1 = 0$ and $c_0 = 1$ and $m = m(p)$. Plugging these two formulae into the definition of $\mathcal{A}_{0,q,\tilde{\mu}}$ in Eq. 1.9, we obtain

$$\begin{aligned}\mathcal{A}_{0,q,\tilde{\mu}}c_n &= \tilde{\mu} \sum_k [q(k-m+1)^n - (k-m)^n - (1-q)(k-m-1)^n] p_k \\ &\quad - \tilde{\mu}(2q-1)nc_{n-1} + \frac{1}{2}n(n-1)c_2c_{n-2} - nc_n \\ &= \frac{1}{2}n(n-1)c_2c_{n-2} - nc_n + \tilde{\mu} \sum_{i=2}^n \binom{k}{i} (q+(-1)^i(1-q))c_{n-i}.\end{aligned}$$

In particular,

$$\begin{aligned}\mathcal{A}_{0,q,\tilde{\mu}}c_2 &= \tilde{\mu} - c_2 \\ \mathcal{A}_{0,q,\tilde{\mu}}c_3 &= \tilde{\mu}(2q-1) - 3c_3 \\ \mathcal{A}_{0,q,\tilde{\mu}}c_4 &= \tilde{\mu}(6c_2+1) + 6c_2^2 - 4c_4 \\ \mathcal{A}_{0,q,\tilde{\mu}}c_5 &= \tilde{\mu}(10c_3+10(2q-1)c_2+(2q-1)) + 10c_3c_2 - 5c_5,\end{aligned}\quad (1.21)$$

and Itô's formula yields

$$\begin{aligned}c_n(P(t)) &= c_n(P(0)) + \int_0^t \mathcal{A}_{0,q,\tilde{\mu}}c_n(P(s)) ds \\ &\quad + \int_0^t \sum_{k,l} [(k-m(P(s))^n - knc_{n-1}(P(s))] \sqrt{P_k(s)P_l(s)} dW_{kl}(s),\end{aligned}\quad (1.22)$$

where the last integral is a martingale. For c_2 , we take expectation on both sides of Eq. 1.22 to obtain

$$E^{\mathbb{P}^0}[c_2(P(t))] = E^{\mathbb{P}^0}[c_2(P(0))] + \int_0^t (\tilde{\mu} - E^{\mathbb{P}^0}[c_2(P(s))]) ds.$$

In differential form, we can write

$$\frac{d}{dt} E^{\mathbb{P}^0}[c_2(P(t))] = \tilde{\mu} - E^{\mathbb{P}^0}[c_2(P(t))].$$

Thus, $E^{\mathbb{P}_0}[c_2(P(t))] = E^{\mathbb{P}_0}[c_2(P(0))]e^{-t} + \tilde{\mu}(1 - e^{-t})$, and hence

$$\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_2(P(t))] = \tilde{\mu}. \quad (1.23)$$

Similarly,

$$\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_3(P(t))] = \frac{\tilde{\mu}}{3}(2q - 1). \quad (1.24)$$

For $n = 4$ and $n = 5$, $\mathcal{A}_{0,q,\tilde{\mu}}c_n$ involves terms of the form $c_{n-2}c_2$ in addition to c_n :

$$\begin{aligned} \mathcal{A}_{0,q,\tilde{\mu}}(c_2^2) &= 2\tilde{\mu}c_2 + c_4 - 3c_2^2, \\ \mathcal{A}_{0,q,\tilde{\mu}}(c_3c_2) &= \tilde{\mu}[c_3 + (2q - 1)c_2] + c_5 - 6c_3c_2. \end{aligned}$$

Combining the above two formulae with Eq. 1.21 yields

$$\begin{aligned} \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_4(P(t))] &= 5\tilde{\mu} \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_2(P(t))] + \frac{\tilde{\mu}}{2} = 5\tilde{\mu}^2 + \frac{\tilde{\mu}}{2}, \\ \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_2^2(P(t))] &= \frac{7}{3} \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_2(P(t))] + \frac{\tilde{\mu}}{2} = \frac{7}{3}\tilde{\mu}^2 + \frac{\tilde{\mu}}{2}, \\ \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_5(P(t))] &= \tilde{\mu}(2q - 1) \left(\frac{5\tilde{\mu}}{3} + \frac{3}{20} \right), \\ \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_3(P(t))c_2(P(t))] &= \tilde{\mu}(2q - 1) \left(\frac{\tilde{\mu}}{2} + \frac{1}{40} \right). \end{aligned} \quad (1.25)$$

Lastly, we need the following result

$$\begin{aligned} d[M, c_n] &= \sum_{k,l,k',l'} ((k-m)^n - nkc_{n-1})k' \sqrt{P_k P_l P_{k'} P_{l'}} d[W_{kl}, W_{k'l'}] \\ &= \left(\sum_k k((k-m)^n - nkc_{n-1})P_k - \sum_{k,l} ((k-m)^n - nkc_{n-1})lP_k P_l \right) dt \\ &= \sum_k (k-m)((k-m)^n - nkc_{n-1})P_k dt \\ &= (c_{n+1} - nc_{n-1}c_2) dt. \end{aligned} \quad (1.26)$$

1.4.3 Representation Using the Neutral Process

From Eq. 1.8, the mean of the selected process \tilde{P} satisfies

$$\begin{aligned} m(\tilde{P}(t)) &= m(\tilde{P}(0)) + \int_0^t \sum_k k[\tilde{\mu}(q\tilde{P}_{k-1}(s) - \tilde{P}_k(s) + (1-q)\tilde{P}_{k+1}(s)) \\ &\quad + \tilde{\sigma}(k - m(\tilde{P}(s)))\tilde{P}_k(s)] ds + M(t) \\ &= \tilde{\mu}(2q - 1)t + \tilde{\sigma} \int_0^t c_2(\tilde{P}(s)) ds + M(t). \end{aligned} \quad (1.27)$$

This is similar to Eq. 1.4, i.e. the strong selection case of model A, where the rate of adaptation is also proportional to the variance of fitness. The critical quantity of interest is the integrated variance $\int_0^t c_2(\tilde{P}(s)) ds$. But if we try to write down the evolution equation of c_2 , we will again obtain a system resembling Eq. 1.12.

For the weak selection model, however, we have a representation of the selected process $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$ in terms of the neutral process, and we can write central moments of the selected process using this representation. For example, let $P(t)$ be the solution to the neutral SDE (Eq. 1.15), then the integrated variance of the selected process associated with $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$ can be written as

$$E^{\mathbb{P}_{\tilde{\sigma}}} \left[\int_0^t c_2(\tilde{P}(s)) ds \right] = E^{\mathbb{P}_0} \left[Z(t) \int_0^t c_2(P(s)) ds \right]. \quad (1.28)$$

We define

$$N(t) = \int_0^t c_2(P(s)) ds$$

and write

$$Z(t) = \exp \left(\tilde{\sigma} M(t) - \frac{\tilde{\sigma}^2}{2} N(t) \right).$$

We expand Z into a Taylor series around 0 and write the term inside the expectation on the right-hand side of Eq. 1.28 as follows, dropping the dependence on t of M and N in our notation:

$$Z(t)N = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\tilde{\sigma} M - \frac{\tilde{\sigma}^2}{2} N \right)^k N.$$

We can write the left-hand side of Eq. 1.28 as an expansion in the selection coefficient $\tilde{\sigma}$:

$$\begin{aligned} E^{\mathbb{P}_{\tilde{\sigma}}} \left[\int_0^t c_2(\tilde{P}(s)) ds \right] &= E^{\mathbb{P}_0} \left[\sum_{k=0}^{\infty} \frac{1}{k!} \left(\tilde{\sigma} M - \frac{\tilde{\sigma}^2}{2} N \right)^k N \right] \\ &= E^{\mathbb{P}_0} \left[N + \left(\tilde{\sigma} M - \frac{\tilde{\sigma}^2}{2} N \right) N + \frac{1}{2} \left(\tilde{\sigma} M - \frac{\tilde{\sigma}^2}{2} N \right)^2 N \right. \\ &\quad \left. + \frac{1}{6} \left(\tilde{\sigma} M - \frac{\tilde{\sigma}^2}{2} N \right)^3 N + \dots \right] \\ &= E^{\mathbb{P}_0} [N] + \tilde{\sigma} E^{\mathbb{P}_0} [MN] + \frac{\tilde{\sigma}^2}{2} E^{\mathbb{P}_0} [(M^2 - N)N] \\ &\quad + \frac{\tilde{\sigma}^3}{6} E^{\mathbb{P}_0} [(M^3 - 3MN)N] + \dots, \end{aligned}$$

where in the last step we have taken the expectation inside the infinite sum, which needs to be justified rigorously. Assuming we can do this, we can quite easily compute the approximate value of the first few terms in the above expansion by hand for large time and they yield some interesting information about the rate of adaptation of the selected process $\mathcal{A}_{\tilde{\sigma}, q, \tilde{\mu}}$. For the first term in the expansion

$$E^{\mathbb{P}_0}[N] = E^{\mathbb{P}_0} \left[\int_0^t c_2(P(s)) ds \right] = \int_0^t E^{\mathbb{P}_0}[c_2(P(s))] ds,$$

since $\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_2(P(t))] = \tilde{\mu}$ by (1.23), N increases at rate $\tilde{\mu}$ for large t .

Now we compute the second term in the expansion $E^{\mathbb{P}_0}[MN]$.

$$d(MN) = M dN + d\text{martingale} = M c_2 ds + d\text{martingale}.$$

Thus, as $t \rightarrow \infty$, $E^{\mathbb{P}_0}[M(t)N(t)]$ increases at a rate of $\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[M(t)c_2(t)]$.

$$\begin{aligned} d(Mc_2) &= M dc_2 + d[M, c_2] + d\text{martingale} \\ &= (-Mc_2 + c_3) dt + d\text{martingale} \end{aligned}$$

by Eq. 1.26. Taking expectation on both sides, we get

$$E^{\mathbb{P}_0}[M(t)c_2(P(t))] = \int_0^t E^{\mathbb{P}_0}[c_3(P(s))] - E^{\mathbb{P}_0}[M(s)c_2(P(s))] ds.$$

By an argument similar to that leading to Eq. 1.23, we should expect

$$\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[M(t)c_2(P(t))] = \lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_3(P(t))] = \frac{\tilde{\mu}}{3}(2q-1),$$

by Eq. 1.24. Thus, $E^{\mathbb{P}_0}[MN]$ increases at a rate of $\frac{\tilde{\mu}}{3}(2q-1)$ as $t \rightarrow \infty$.

For the third term in the expansion $E^{\mathbb{P}_0}[(M^2 - N)N]$, we first observe that $M(t)^2 - N(t)$ is a martingale and can be written as $\int_0^t 2M(s) dM(s)$. Therefore

$$\begin{aligned} d((M^2 - N)c_2) &= (M^2 - N) dc_2 + d[M^2 - N, c_2] + d\text{martingale} \\ &= -(M^2 - N)c_2 dt + 2Md[M, c_2] + d\text{martingale} \\ &= [-(M^2 - N)c_2 + 2Mc_3] dt + d\text{martingale}. \end{aligned}$$

by Eq. 1.26. So we need to calculate $\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[M(t)c_3(P(t))]$:

$$\begin{aligned} d(Mc_3) &= M dc_3 + d[M, c_3] + d\text{martingale} \\ &= (-3Mc_3 + c_4 - 3c_2^2) dt + d\text{martingale}, \end{aligned}$$

by Eqs. 1.21 and 1.26. From Eq. 1.25, we have

$$\lim_{t \rightarrow \infty} E^{\mathbb{P}_0}[c_4(P(t)) - 3c_2(P(t))^2] = -2\tilde{\mu}^2 - \tilde{\mu},$$

which means that

$$\begin{aligned} \lim_{t \rightarrow \infty} E^{\mathbb{P}^0} [(M(t)^2 - N(t))c_2(P(t))] &= 2 \lim_{t \rightarrow \infty} E^{\mathbb{P}^0} [M(t)c_3(P(t))] \\ &= \frac{2}{3} \lim_{t \rightarrow \infty} E^{\mathbb{P}^0} [c_4(P(t)) - 3c_2(P(t))^2] \\ &= -\frac{2}{3}(2\tilde{\mu}^2 + \tilde{\mu}). \end{aligned}$$

So we should expect $E^{\mathbb{P}^0} [(M^2 - N)N]$ to increase at a rate of $-\frac{2}{3}(2\tilde{\mu}^2 + \tilde{\mu})$ as $t \rightarrow \infty$.

Similarly, we can calculate that $E^{\mathbb{P}^0} [(M^3 - 3MN)N]$ increases at a rate of $4\tilde{\mu}(2q-1)(\frac{\tilde{\mu}}{6} + \frac{1}{40})$ as $t \rightarrow \infty$. Thus, we can write down the first few terms of the expansion of the left-hand side of Eq. 1.28 for large t

$$\tilde{\mu}t + \tilde{\sigma} \frac{\tilde{\mu}}{3}(2q-1)t - \tilde{\sigma}^2 \frac{1}{3}(2\tilde{\mu}^2 + \tilde{\mu})t + \tilde{\sigma}^3 \frac{2}{3}(2q-1) \left(\frac{\tilde{\mu}^2}{6} + \frac{\tilde{\mu}}{40} \right) t + \dots$$

Plugging this result into Eq. 1.27, we see that we can write the limiting adaptation rate of the process (Eq. 1.8) as an expansion in $\tilde{\sigma}$, where the first few terms are as follows:

$$\begin{aligned} \tilde{\mu}(2q-1) + \tilde{\sigma}\tilde{\mu} + \tilde{\sigma}^2(2q-1)\frac{\tilde{\mu}}{3} - \tilde{\sigma}^3 \left(\frac{2}{3}\tilde{\mu}^2 + \frac{\tilde{\mu}}{3} \right) \\ + \tilde{\sigma}^4(2q-1) \left(\frac{\tilde{\mu}^2}{9} + \frac{\tilde{\mu}}{60} \right) + \dots \end{aligned} \quad (1.29)$$

The first term above is simply the adaptation rate for the neutral process $\mathcal{A}_{0,q,\tilde{\mu}}$. The second term can be interpreted as the first-order correction on the adaptation rate assuming the selected process $\mathcal{A}_{\tilde{\sigma},q,\tilde{\mu}}$ has the same variance of fitness $\tilde{\mu}$ as the neutral process $\mathcal{A}_{0,q,\tilde{\mu}}$. We expect the selected process to have a smaller variance of fitness than the neutral process. But the second-order correction (the $\tilde{\sigma}^2$ term) actually increases the variance if $2q-1 > 0$. If $2q-1 < 0$, on the other hand, then all correction terms on the variance of fitness are negative, as expected. It is the third-order correction (the $\tilde{\sigma}^3$ term) that starts to reduce the variance of fitness in the case of $2q-1 > 0$.

1.5 Open Problems

Although we have laid out in Sect. 1.4.3 a way to write the adaptation rate in the form of a series in the selection coefficient $\tilde{\sigma}$, and demonstrated how, in principle, one can calculate every term in this series, any term above $\tilde{\sigma}^5$ becomes very tedious to calculate by hand. For the series to be of use for realistic ranges of $\tilde{\sigma}$, say 10–100, an efficient way of calculating higher-order terms in the series (Eq. 1.29)

becomes necessary. The selection coefficient $\tilde{\sigma}$ in the weak selection model roughly corresponds with σN in the strong selection model. As N becomes larger, the strong selection model becomes a better approximation. But how large does N need to get for the strong selection model to be a reasonable approximation? From simulations, $\sigma N \geq 500$ seems to be sufficient. Is there a reasonable way to quantitatively measure how good or bad the strong selection model is, so that we obtain some sort of confidence interval for the prediction on adaptation rate we obtain using this model?

Acknowledgements The authors are grateful to Nick Barton for posing the question of whether there exists an upper limit to the adaptation rate in the asexual case. We are also grateful to Nick Barton and Jonathan Coe for valuable discussions throughout this research. In addition, we thank Charles Cuthbertson as a co-author of Cuthbertson et al. (2007). F.Y. was supported by EP-SRC/GR/T19537 while at the University of Oxford.

References

- Barton NH, Coe JB (2007) An upper limit to the rate of adaptation. Preprint
- Cuthbertson C, Etheridge AM, Yu F (2007) Asymptotic behaviour of the rate of adaptation. arXiv:0708.3453
- Dawson DA (1993) Measure-valued Markov processes. In: *École d'Été de Probabilités de Saint-Flour XXI—1991*. Lecture notes in mathematics, vol 1541. Springer, Berlin, pp 1–260
- Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176:1759–1798
- Donnelly P, Kurtz TG (1999) Genealogical processes for Fleming–Viot models with selection and recombination. *Ann Appl Probab* 9(4):1091–1148
- Etheridge AM, Pfaffelhuber P, Wakolbinger A (2007) How often does the ratchet click? Facts, heuristics, asymptotics. *Trends Stochastic Anal.* arXiv:0709.2775v1
- Etheridge AM (2000) An introduction to superprocesses. University lecture series, vol 20. American Mathematical Society, Providence
- Fisher RA (1930) *The genetical theory of natural selection*. Clarendon, Oxford
- Gerrish PJ, Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102/103:127–144
- Gillespie JH (1991) *The causes of molecular evolution*. Oxford University Press, Oxford
- Gordo I, Charlesworth B (2000) On the speed of Muller's ratchet. *Genetics* 156:2137–2140
- Haigh J (1978) The accumulation of deleterious genes in a population. *Theor Popul Biol* 14(2):251–267
- Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc Camb Philos Soc* 23:834–844
- Hegreness M, Shores N, Hartl D, Kishony R (2006) An equivalence principle for the incorporation of favourable mutations in asexual populations. *Science* 311:1615–1617
- Higgs P, Woodcock G (1995) The accumulation of mutations in asexual populations, and the structure of genealogical trees in the presence of selection. *J Math Biol* 33:677–702
- Ikeda N, Watanabe S (1981) *Stochastic differential equations and diffusion processes*. North-Holland mathematics library. North-Holland, Amsterdam
- Karatzas I, Shreve SE (1991) *Brownian motion and stochastic calculus*, 2nd edn. Graduate texts in mathematics, vol 113. Springer, New York
- Novella IS, Elena SF, Moya A, Domingo E, Holland JJ (1995) Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J Virol* 69:2869–2872

- Novella IS, Elena SF, Moya A, Domingo E, Holland JJ (1999) Exponential fitness gains of RNA virus populations are limited by bottleneck effects. *J Virol* 73:1668–1671
- Rouzine I, Brunet E, Wilke CO (2007) The traveling wave approach to asexual evolution: Muller’s ratchet and speed of adaptation. arXiv:0707.3469
- Rouzine I, Wakeley J, Coffin JM (2003) The solitary wave of asexual evolution. *Proc Natl Acad Sci USA* 100(2):587–592
- Shiga T (1982) Wandering phenomena in infinite-allelic diffusion models. *Adv Appl Probab* 14(3):457–483
- Stephan W, Chao L, Smale J (1993) The advance of Muller’s ratchet in a haploid asexual population: approximate solution based on diffusion theory. *Genet Res* 61:225–232
- Wilke CO (2004) The speed of adaptation in large asexual populations. *Genetics* 167:2045–2054