Collective Fluctuations in the Dynamics of Adaptation and Other Traveling Waves

Oskar Hallatschek^{*,1} and Lukas Geyrhofer[†]

^{*}Biophysics and Evolutionary Dynamics Group, Departments of Physics and Integrative Biology, University of California, Berkeley, California 94720-3220, and [†]Biophysics and Evolutionary Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, 33077 Göttingen, Germany

ABSTRACT The dynamics of adaptation are difficult to predict because it is highly stochastic even in large populations. The uncertainty emerges from random genetic drift arising in a vanguard of particularly fit individuals of the population. Several approaches have been developed to analyze the crucial role of genetic drift on the *expected* dynamics of adaptation, including the mean fitness of the entire population, or the fate of newly arising beneficial deleterious mutations. However, little is known about how genetic drift causes fluctuations to emerge on the population level, where it becomes palpable as variations in the adaptation speed and the fitness distribution. Yet these phenomena control the decay of genetic diversity and variability in evolution experiments and are key to a truly predictive understanding of evolutionary processes. Here, we show that correlations induced by these emergent fluctuations can be computed at any arbitrary order by a suitable choice of a dynamical constraint. The resulting linear equations exhibit fluctuation-induced terms that amplify short-distance correlations and suppress long-distance ones. These terms, which are in general not small, control the decay of genetic diversity and, for wave-tip dominated (“pulled”) waves, lead to anticorrelations between the tip of the wave and the lagging bulk of the population. While it is natural to consider the process of adaptation as a branching random walk in fitness space subject to a constraint (due to finite resources), we show that other traveling wave phenomena in ecology and evolution likewise fall into this class of constrained branching random walks. Our methods, therefore, provide a systematic approach toward analyzing fluctuations in a wide range of population biological processes, such as adaptation, genetic meltdown, species invasions, or epidemics.

KEYWORDS traveling-wave models; asexual adaptation; Fisher waves; tuned models; stochastic modeling; higher-order correlation functions; exact approaches

MANY important evolutionary and ecological processes rely on the behavior of a small number of individuals that have a large dynamical influence on the population as a whole. This is, perhaps, most obvious in the case of biological adaptation in asexual species, such as microbes (over time-scales where horizontal gene transfer is irrelevant). Future generations descend from a small number of currently well-adapted individuals. The genetic footprint of the large majority of the population is wiped out over time by the fixation of more fit genotypes. For a large influx of beneficial mutations, these dynamics can be visualized as a traveling wave in fitness

space; see Figure 1A. At any time, the currently most fit “pioneer” individuals reside in the small tip of the wave. As time elapses, the wave moves toward higher fitness and the formerly rare most fit individuals dominate the population. By that time, however, a new wave tip of even more fit mutants has formed by mutations and the cycle of transient dominance continues. Evolutionary dynamics in sexual species differ in the sense that recombination gives individuals a chance to inherit well-adapted genes even if they do not belong to the class of most fit individuals. Nevertheless, even in sexual species, wave-like dynamics can arise in chromosomal fragments, within a single locus (Neher *et al.* 2013; Weissman and Hallatschek 2014).

The principle of “a few guiding the way for many” also characterizes the motion of flocks of birds, which can be controlled by just a few leaders, as well as the expansion of an

Copyright © 2016 by the Genetics Society of America
doi: 10.1534/genetics.115.181271

Manuscript received August 2, 2015; accepted for publication January 13, 2016; published Early Online January 27, 2016.

¹Corresponding author: Physics and Integrative Biology, University of California, Stanley Hall No. 274D, Berkeley, CA 94720-3220. E-mail: ohallats@berkeley.edu

Figure 1 Models of adaptation and invasion. An illustration of the essential features of two types of noisy traveling waves is shown. (A) Well-mixed models of asexual adaptation generate waves that travel across fitness landscapes toward higher fitness. (B) Models of species invasions, range expansions, or epidemics generate waves that propagate in real space. Both types of waves can be viewed as emerging from constrained branching random walks. The state of the traveling wave is described by a (fluctuating) population density shown in blue. A branching process is generated by linear net growth rates and a jump process. (C) In the case of adaptive waves, individuals reproduce according to their current fitness and jump due to spontaneous mutations. (D) By contrast, the jump process in invasion waves is generated by random movements according to some dispersal kernel, and density-dependent growth rates lead to an effective location dependence of the growth rates. A global constraint ensures a finite population size and depends on a weighting function $u_t(x)$ indicated in green. Note that for conventional models of adaptation with fixed population size, the growth rates are increasing with fitness and the weighting function is a constant. By contrast, for models of species invasion or range expansion without the Allee effect (Hallatschek and Nelson 2008), the growth rates and weighting function have a sigmoidal shape, saturating in the tip of the wave. This ensures that individuals in the tip of the wave have the highest growth rates (because of the least competition) and that their total number stays finite.

invasive species, which depends on pioneers most advanced into the virgin territory. In these cases, one obtains traveling waves in real space rather than in fitness space, as is illustrated in Figure 1B.

The resulting overall dynamics in both fitness and real-space waves can become highly erratic even in large populations because the behavior of the entire population is influenced by strong number fluctuations, called genetic drift, occurring in the small subset of pioneer individuals. Such propagation processes with an extreme sensitivity of noise have also been called “pulled” waves, because they are pulled along by the action of the most advanced individuals (van Saarloos 2003). In “pushed” waves (Barton 1979; van Saarloos 2003), by contrast, most of the growth occurs behind the front at higher population densities. While these pushed waves allow for simple mean-field approximation that neglects noise, pulled waves break down when noise is neglected. The reason is that noise is a singular perturbation and neglecting it can lead to qualitatively wrong predictions or even divergences.

If one ignores the fluctuations at the population level and is interested only in the expected dynamics of the population, one might be tempted to simply ignore genetic drift in models of pulled waves. However, it turns out that mean-field models ignoring genetic drift drastically overestimate the speed of traveling waves, to the point that they predict an ever accelerating rather than a finite speed of adaptation. It took 70 years since the first formulation of traveling wave models by

Fisher and Kolomogorov to clarify that genetic drift influences both the expectation and the variation of pulled waves in a singular manner (Tsimring *et al.* 1996; Brunet and Derrida 1997): Genetic drift is a correction that cannot be treated by a perturbation analysis.

The *expected* behavior of pulled waves has since been studied at great length. Most theoretical approaches describe the state of a population by a distribution over trait values on which evolutionary forces are acting. Such models are equivalent to “continuum of alleles” models, which have been used extensively to study deterministic population genetic equilibria, such as balancing selection (Kimura 1965; Lande 1975; Fleming 1979; Turelli 1984; Johnson and Barton 2005). The key challenge was then to make analytical progress in applying this framework to problems that do not have a well-defined deterministic limit. Many results were first obtained for waves of invasion, noisy versions of the classical “FKPP” model by Fisher, Kolmogorov, Petrovskii, and Piskunov (van Saarloos 2003). In recent years, however, there has been a particularly strong research focus on models of adaptation. These models aroused widespread interest because they can be applied to several types of data, including genomic data derived from experimental evolution experiments and from natural populations that undergo rampant adaptation, such as bacteria and viruses (Neher 2013). We now have analytical or semianalytical results for a number of observables such as the mean speed, probability of fixation, and distribution of

fixed mutations in the asymptotic regime of large populations (Neher 2013). The great value of these results is that they show and rationalize which parameter combination chiefly influences the dynamics and through which functional form.

Models of asexual adaptation generally exhibit two regimes with drastically different dynamics. For small population sizes or low mutation rates, beneficial mutations enter the population only very rarely. The speed of adaptation then depends only on the product of the rate μN at which beneficial mutations enter the population and the probability $2s$ at which they survive despite genetic drift. This scenario of periodic selection was long understood, but it failed to describe many microbial evolution experiments in which adaptation is rapid (Kawecki *et al.* 2012). The problem is that the waiting time between two establishing beneficial mutants is typically less than the time to fixation in microbial evolution experiments.

Therefore, it was important to understand how the evolutionary dynamics for large population sizes are controlled by the competition between many beneficial mutations arising on different genetic backgrounds. This phenomenon of clonal interference causes a diminishing return in population size and mutation rate. As one increases either of these parameters, one increases linearly the number of occurring competing mutations. The rate of fixing beneficial mutations, however, is changed to a much lower degree. As a consequence, the dynamics of adaptation have been found to depend only logarithmically on population size and mutation rate. This weak functional dependence on the influx of beneficial mutations in the dynamics of clonal interference is in fact at the root of universality observed in many such wave models: These predictions are independent of the precise details of the models, including the form of the nonlinear population size regulation.

The basic challenge in analyzing models of adaptation, as well as other pulled waves, arises from an essential nonlinearity that is required to control population growth. Ignoring such a dynamical control of the population size leads to long-term exponential growth or population extinction. Progress in describing the *mean* behavior of front-sensitive models has been achieved by at least three different approaches: One can heuristically improve the mean-field dynamics by setting the net growth rate equal to zero in regions where the population densities are too small (Tsimring *et al.* 1996). Such an *ad hoc* approach, based on a growth rate cutoff, correctly reproduces the wave speed to the leading order but does not reveal other universal next-to leading-order corrections. One can also invoke a branching-process approximation for the tip of the wave, thereby neglecting effects of the nonlinear population size control, and then match this linearized description with a deterministic description of the bulk of the population (Rouzine *et al.* 2003, 2008; Desai and Fisher 2007; Schifflers *et al.* 2011; Good *et al.* 2012). Finally, there is also the possibility to invoke a particular dynamical constraint with the property that the dynamics exhibit a closed linear equation for the first moment. Importantly, this method, which has been called

“model tuning” (Hallatschek 2011; Good *et al.* 2012; Geyrhofer and Hallatschek 2013), reproduces the universal features in the large population size limit, which are independent of the chosen population control, ultimately because of the weakness of the population size dependence. Since this approach also reproduces the correct behavior in the small population size limit when mutational effect sizes are small, it may be viewed as a practical interpolation scheme for the entire range of population sizes (Hallatschek 2011). Nevertheless, the model-tuning approach is most appropriate for studying populations large enough such that the population density in fitness space can be represented in the form of a wave.

While understanding the mean behavior of noisy traveling waves has been an important achievement, the actual stochastic dynamics are characterized by pronounced fluctuations at the population level. No two realizations of an evolution experiment, for instance, will exhibit the same time-dependent fitness distribution because of the chance effects involved in reproduction and mutations. Measuring the mean behavior requires many replicates in which the entire environment is accurately reproduced. Even if one has access to many replicates, as is possible in highly parallelized well-mixed evolution experiments, one can potentially learn a lot from the variability between replicates. Thus, a predictive understanding of the variability in evolutionary trajectories would greatly improve our quantitative understanding of how evolution works.

Moreover, fluctuations in the shape of traveling waves are also crucial for understanding the genetic diversity within an adapting population. A classic strategy to visualize the internal dynamics in evolution experiments is to start with a mixture of strains that carry different genetic or fluorescent markers and then monitor the marker ratio over time, both in well-mixed (Hegreness *et al.* 2006; Kawecki *et al.* 2012; Barroso-Batista *et al.* 2014) and in spatial eco-evolutionary experiments (Hallatschek *et al.* 2007; Müller *et al.* 2014; Weber *et al.* 2014). The loss of marker diversity can be a measure for the emergence of competing clones in well-mixed populations or of allele surfing events in expanding populations. Predicting the decay of marker diversity reflects fluctuations in the composition and shape of the population because, on average, the ratio of neutral markers stays constant.

A few results on fluctuations at the population level are available (Brunet *et al.* 2006; Hallatschek and Korolev 2009; Neher and Shraiman 2012; Fisher 2013; Good and Desai 2013), focusing on the small or large speed regimes in models of adaptation and of background selection. However, we currently still lack a systematic approach that can be applied to a wide range of models. Here, we fill this gap by extending the method of model tuning (Hallatschek 2011) to the analysis of higher correlation functions: We show that it is possible to choose a dynamical constraint in such a way that the hierarchy of moments is closed at any desired order. The resulting linear equations have intuitive interpretations, can be solved numerically, and are amenable to asymptotic analytical techniques. The overall framework manifests itself in a

particular limiting mathematical structure, which we call *constrained branching random walks*. In applying this approach, we first focus on models of asexual adaptation and show specifically how correlations and the gradual decay of genetic (or marker) diversity can be computed. We also demonstrate that our approach can be applied to study genetic diversity in range expansions. In addition, we can envision a wide range of models in ecology and evolution for which our framework provides a capable tool to obtain results on fluctuation properties of the system.

Models of Adaptation as Constrained Branching Random Walks

Darwinian adaptation emerges from the processes of mutation, reproduction, and competition, and these features need to be mirrored in any model of adaptation. In models, spontaneous mutations can be represented by a stochastic jump process in a “fitness space.” Reproduction is naturally described by a branching process by which individuals give birth at certain fitness-dependent rates (Allen 2003; Haccou *et al.* 2005). In combination, reproduction and mutations thus generate a branching random walk (Allen 2003; Haccou *et al.* 2005), which by itself would lead to diverging population sizes. To avoid this unrealistic outcome, models of adaptation also encode a constraint on population sizes to account for the competition for finite resources. The resulting process is a branching random walk subject to a global constraint, which we now frame mathematically.

The state of the population at time t is described by a function $c_t(x)$ representing the number density of individuals with fitness x . In this context, fitness refers to an individual's net growth rate in the absence of competition for resources. The population is assumed to evolve in discrete time steps of size ε , which is eventually set to zero to obtain a continuous-time Markov process. Each time step consists of two substeps. The first substep realizes reproduction and mutations and the second substep implements competition.

First substep: Reproduction and mutations

The combined effect of reproduction and mutations can be described by the stochastic equation

$$\tilde{c}_{t+\varepsilon} - c_t = \varepsilon \mathcal{L}_t c_t + \sqrt{\varepsilon b c_t} \eta_t, \quad (1)$$

which takes the number density c_t to an intermediate value $\tilde{c}_{t+\varepsilon}$. The term $\varepsilon \mathcal{L}_t c_t$ represents the expected change in density due to reproduction and mutations. This term is linear in the number density because the number of offspring and mutants per time step is proportional to the current population density. The term $\sqrt{\varepsilon b c_t} \eta_t$ represents all sources of noise arising in this setup. We now discuss separately the precise meaning of both terms and give natural alternatives for their form.

The mutation–selection operator $\mathcal{L}_t$ can be cast into the mathematical form

$$\mathcal{L}_t = \mathcal{M}_t + x - x_0(t). \quad (2)$$

The linear “reaction” term $x - x_0$ in $\mathcal{L}_t$ simply accounts for the fact that individuals with higher growth rate x grow faster. The term $x_0(t)$ refers to the mean fitness of the population, which separates the population with a positive net growth rate $x > x_0$ from the less fit part of the population with $x < x_0$.

The mutational process is described by the operator $\mathcal{M}_t$, which conserves particle numbers, *i.e.*, describes a pure jump process. Common choices of increasing complexity for the mutational process are

$$\{\mathcal{M}_t c_t\}(x) \sim \begin{cases} D \partial_x^2 c_t(x), \\ \mu [c_t(x-s) - c_t(x)], \\ \int_y \mu(y) [c_t(x-y) - c_t(y)]. \end{cases} \quad (3)$$

From top to bottom, these expressions correspond to a “diffusion kernel,” to a “staircase kernel,” and in the last line to a “general mutational kernel.” Here μ is a mutation rate and s is a characteristic scale for the mutational effect. The diffusion kernel has been frequently employed in the population genetics literature (Kimura 1965; Lande 1975; Fleming 1979; Turelli 1984; Tsimring *et al.* 1996; Johnson and Barton 2005; Hallatschek 2011) because it is the simplest of these kernels: It is characterized by only one compound parameter, the diffusion constant $D = \int_y y^2 \mu(y)/2$, rather than two in the staircase kernel or an entire function $\mu(y)$ in the general case. The diffusion constant D quantifies the fitness variance per generation, generated by an influx of novel mutations.

One cannot generally assume that (biased) diffusion is a good model for discrete mutational events because of the presence of the reaction term favoring highly fit individuals. The diffusion kernel assumes that fitness is controlled by many mutations of sufficiently weak effect (Turelli 1984) such that mutation rates are higher than the typical fitness effects of novel mutations. This may apply to rapidly mutating organisms, such as viruses (Tsimring *et al.* 1996; Hallatschek 2011), or be close to a dynamic mutation–selection balance (Goyal *et al.* 2012; Neher *et al.* 2013). It may also effectively apply in island models with low migration rates, where the fitness effect of a mutation is reduced by potentially low migration rates. But, in well-mixed populations, the diffusion approach breaks down when the probability per generation to acquire a beneficial mutation is much smaller than their typical effect (Rouzine *et al.* 2003), which has been confirmed for a number of microbial species when they adapt to new environments (Perfeito *et al.* 2007; Gordo *et al.* 2011; Levy *et al.* 2015).

The stochastic term $\sqrt{\varepsilon b c_t} \eta_t$ in Equation 1 accounts for all random factors that influence the reproduction process. The function $\eta_t(x)$ represents standard white noise, *i.e.*, a set of δ -correlated random numbers,

$$\overline{\eta_t(x) \eta_{t'}(y)} = \delta(x-y) \delta_{tt'}, \quad (4)$$

where $\bar{\cdot}$ denotes the ensemble average of a random variable f , and $\delta(x)$ and δ_{ij} are the Dirac delta function and the

Kronecker delta, respectively. The amplitude $\propto \sqrt{\epsilon b c_t}$ of the noise term in Equation 1 is typical for number fluctuations: Due to the law of large numbers, the expected variance in population numbers from one time step to the next is proportional to the number ϵc_t of expected births or deaths during one time step. An important effect of the noise term is to make the fate of newly arising mutants uncertain: Most of the mutations get lost due to genetic drift. The probability to survive in the initial phase of drift is inversely proportional to the parameter b . In general, the numerical coefficient b is given by the variance in offspring number per individual. For instance, when we assume that offspring have nearly matching division and death rates, we find $b = 2$; this limit corresponds to the continuous-time branching process. In practice, the variance in offspring number is often assumed to be of order one, unless the offspring distribution is highly skewed, which can occur when few individuals produce most of the offspring (Hedgcock and Pudovkin 2011).

Second substep: Population size constraint

Because the reproduction step generally changes population numbers, another substep, following the branching process, is required to enforce a constant population size,

$$1 = \int_x \frac{1}{N} c_t(x) . \quad (5)$$

[Note that if $x_0(t)$ denotes the *mean fitness* of the population, the action of the Liouvillean $\mathcal{L}$ does not change the expected number of individuals in the population. However, fluctuations in the reproduction implemented by genetic drift will result in slight deviations from this expected outcome. These deviations accumulate over time and lead either to extinction or to an ever increasing population size.] In most models and experiments (Gerrish and Lenski 1998; Desai *et al.* 2007), this step is realized by a random culling of the population: Individuals are eliminated at random from the population until the population size constraint is restored. Mathematically, the population control step can be cast into the form

$$c_{t+\epsilon} = \tilde{c}_{t+\epsilon}(1 - \lambda), \quad (6)$$

where λ represents the fraction of the population that has to be removed to comply with the population size constraint. Note that the culling process does not discriminate among individuals of different fitness.

The second substep completes the computational time step and takes the concentration field from the intermediate state $\tilde{c}_{t+\epsilon}$ to the properly constrained state $c_{t+\epsilon}$.

Generalization to arbitrary linear constraints

While it is necessary to constrain the population dynamics to avoid an exponential runaway, there is no particular biological reason to strictly fix the population size—in fact, most population sizes fluctuate over time (Frankham 1995). As we will see, there are, however, mathematical reasons to consider

constraints of a particular form, which greatly simplify the analysis.

As a key step toward these tuned models, we note the fixed population size constraint Equation 5 can be viewed as one member of a whole class of linear constraints,

$$1 = \int_x u_t(x) c_t(x) \equiv \langle u_t | c_t \rangle, \quad (7)$$

that one could formulate with the help of a suitable weighting function $u_t(x)$. Note that one recovers the fixed population size constraint if one chooses the weighting function $u_t(x)$ to be a constant, $u = N^{-1}$. For any other choice, one obtains a unique model of adaptation defined by the weighting function $u_t(x)$ and the mutation–selection operator $\mathcal{L}$. Our main result is the observation that there is an entire set of weighting functions for which the dynamics of the model become simple (*cf. Summary of Main Formal Results*).

Invasion Waves as Constrained Branching Random Walks

Regulating population densities by adjusting an overall growth rate $\lambda(t)$ is natural when one wants to model microbial evolution experiments that frequently go through rounds of dilution. It is not obvious that such a density regulation also works for other types of pulled waves arising, for instance, in ecology. Therefore, we also test to what extent our theoretical framework applies to traveling waves that are used to describe species invasions or range expansions.

Figure 1B illustrates the expansion of a population in a real landscape, which may describe an advantageous gene spreading through a population distributed in space or the invasion of virgin territory by an introduced species. In contrast with models of adaptation, one obtains a wave train rather than a wave “packet,” and the wave train moves in real space rather than in fitness space. Models of such real-space waves require the following features: (i) Populations reproduce and die “freely” in the tip of the wave, where population densities are small; (ii) individuals move in real space according to some jump process; and (iii) the net growth vanishes in the bulk of the wave, where resources are sparse.

Features i and ii again generate a branching random walk in the tip of the wave, however, according to different position-dependent growth and jump rates than in the case of evolution. Feature iii requires a finite population size in the tip of the wave. This nonlinearity keeps the branching random walk away from proliferating to infinite densities, where mean-field models apply.

To generate an adequate branching random walk, one has to use an appropriate Liouville operator $\mathcal{L}$, which represents the ecological processes of dispersal and density growth. In one dimension, one can choose, for instance,

$$\mathcal{L}_t = \mathcal{M}_t + sH[x - x_0(t)]. \quad (8)$$

Here, x refers to the location in one-dimensional real space and $H(z)$ is the unit step function, defined by $H(z \leq 0) = 1$

and $H(z < 0) = 0$. x_0 refers to the position of the crossover to the bulk of the wave. The operator $\mathcal{M}_t$ generates a jump process. In the simplest case, again, the jump process may be approximated by a diffusive process, $\mathcal{M}_t c = D \partial_x^2 c$. The growth term does not need to be modeled by a strict step function, as is done in Equation 8; any sigmoidal function works in the limit of large population sizes (Brunet and Derrida 1997). Importantly, the net reproduction rates need to rise monotonously with x and saturate at some finite value s at $x = O(1)$. Finally, to limit the population size in the front of the wave, we need to use a nonconstant weighting function $u(x, t)$ in the global constraint, for instance, $u_t(x) = (1/N)H[x - x_0]$, which ensures that the growth region contains precisely N individuals.

The qualitative difference between the traveling waves in real and fitness space turns out to rely on the growth rates in the nose of the wave: While growth rates are saturating in the case of real-space waves, they increase without bound in the case of waves of adaptation. This makes models of adaptation even more sensitive to the effects of noise to the extent that a mean-field limit (neglecting noise) does not even yield finite velocity waves. Noise serves a crucial function in these models as it is required to regularize the wave dynamics. These waves have therefore been called “front-regularized waves” (Cohen *et al.* 2005a).

Summary of Main Formal Results

In the same way as exemplified above for models of adaptation and invasion, one can frame many other eco-evolutionary scenarios, in their essence, as constrained branching random walks. These models are, ultimately, defined by an operator $\mathcal{L}_t$ generating the branching random walk and a global constraint parameterized by a weighting function u_t . In this article, we show that, in fact, for any given $\mathcal{L}_t$ there are *natural* ways of choosing the weighting function. The associated models, which we call *tuned* models, have desirable properties, including closed and linear moment equations, greatly facilitating their analysis.

We first state our main results on how to construct tuned models at any desired order of moments. We then provide an interpretation of these tuned models and simulation results, which we compare to fixed population size models. Detailed analytical derivations are given in *Appendices A–F*.

To characterize fluctuations in the makeup of the population it is convenient to consider the so-called n -point correlation function $\bar{C}^{(n)}$, which is the noise average of the product

$$C^{(n)}(x_1, \dots, x_n; t) \equiv \prod_{j=1}^n c(x_j; t), \quad (9)$$

of n number density fields, $c(x_j; t)$, evaluated at the same time t at various locations x_j . Note that, here and henceforth, we use the notations $f_t(x)$ and $f(x; t)$ for a space- and time-dependent function interchangeably.

From many studies over the last 15 years, we know a lot about the first moment, $\bar{C}^{(1)}(x; t)$, of several models of noisy traveling waves. This provides access to the expected shape and velocity of the traveling wave, as well as to the expected fate of individual mutations or subpopulations. However, wave shape and velocity fluctuations, as well as the decay of genetic diversity, require access to higher moments, $n > 1$, for which there is no systematic approach so far.

Our main result is that the dynamics of the n th moment (and of all lower moments) become analytically accessible if one chooses in the global constraint Equation 7 the weighting function u_t to have the form

$$u^{(n)}(x; t) \equiv \frac{2w(x; t)}{b(n+1)} \quad (10)$$

for any positive integer n , where $w(x; t)$ satisfies

$$-\partial_t w = [\mathcal{L}_t^\dagger + \gamma(t) - w]w \quad (11)$$

with the adjoint operator $\mathcal{L}_t^\dagger$ of $\mathcal{L}_t$. The arbitrary function $\gamma(t)$ can be used to control the dynamics of the system. For instance, a temporary negative $\gamma(t)$ decreases overall population size, while $\gamma(t) > 0$ increases it. For all practical examples in this article, we choose $\gamma = 0$ to keep the mean population size constant. [The meaning of the control parameter function $\gamma(t)$ was discussed in Hallatschek 2011, where it was denoted by $x_0(t)$.]

For the special weighting function Equation 10, the equation of motion for the n th moment becomes *closed* and *linear* and is given by

$$\begin{aligned} \partial_t \bar{C}^{(n)} = & \sum_{j=1}^n \left(\mathcal{L}_t + \gamma(t) - \frac{2n}{n+1} w \right) \Big|_{x_j} \bar{C}^{(n)} \\ & + \frac{2}{n+1} \sum_{j=1}^n \sum_{k=j+1}^n \delta(x_j - x_k) \langle w | \bar{C}^{(n)} \rangle_k. \end{aligned} \quad (12)$$

The resulting models may be called *tuned*, because for all other choices of the weighting function, the equation of motion for $\bar{C}^{(n)}$ involves $\bar{C}^{(n+1)}$, resulting in an infinite hierarchy of moments. The moment closure of the tuned models is exact and not due to a truncation or approximation of higher moments. The two terms $\propto w$ in Equation 12 are fluctuation-induced terms, which are absent in the deterministic approximation. The last, positive term generates correlations at equal space arguments, which are then dissipated by the first, strictly negative, term over longer timescales. The positive term exists only for $n > 1$ and indicates that the effect of genetic drift on higher moments is more complicated than a cutoff in the growth term (Tsimring *et al.* 1996; Brunet and Derrida 1997; Cohen *et al.* 2005b).

Moreover, examining the behavior of differently labeled, but otherwise identical, subpopulations shows that tuned models can be interpreted naturally in the framework of population genetics. First, the weighting function of all

tuned models is a fixation probability function: The probability that descendants of an individual at location x and time t will take over the population on long times is $u^{(n)}(x; t) = 2w(x; t)/b(n+1)$ in the model tuned for the n th moment. It is remarkable, in this context, that Equation 11 for $w(x; t)$ is precisely the equation governing the survival probability of an *unconstrained* branching random walk (Fisher 2013), with $b = 2$.

Second, the higher moments provide access to the statistics of the genetic makeup of the population. It is a key insight in population genetics that the genetic diversity of a population must decrease with time in the absence of mutations: Fixation and extinction of subtypes need to be maintained by an appropriate influx of mutations. The decay of genetic diversity fundamentally depends on higher moments—the first moment captures fixation probabilities but not the time to fixation.

A type of diversity that is often not replenished on the timescale of experiments is that of neutral markers. The decay of such a marker diversity can be quantified by the cross-correlation functions (Nagylaki 1978), which are important population genetic observables accessible both in natural populations (Slatkin 1985) and in laboratory evolution experiments (Hegreness *et al.* 2006). Thus, we consider the expression

$$\overline{C}_{i_1, \dots, i_n}^{(n)}(x_1, \dots, x_n; t) \equiv \prod_{j=1}^n c_{i_j}(x_j; t), \quad (13)$$

for (different) subtypes $i_1, \dots, i_n$. Here, the number density $c_{i_j}(x; t)$ is the number density field of type i_j . The sum of all subtypes makes up the total population, $c(x; t) = \sum_i c_i(x; t)$. The case $n = 2$ is the most prominent one: $\overline{C}_{1,2}^{(2)}(x, x; t)$ is proportional to the so-called heterozygosity at location x and t , which is the probability that two individuals sampled with replacement are of different type (Wakeley 2009). If all i_j are different and $n > 1$, it is clear that $\overline{C}_{i_1, \dots, i_n}^{(n)}$ must gradually decay because one of the subtypes will take over in the presence of random genetic drift.

We find that, provided all i_j are different, the cross-correlation function in the n th-tuned model can be expressed in the simple form

$$\overline{C}_{i_1, \dots, i_n}^{(n)} = \prod_{k=1}^n f_{i_k}(x_{i_k}; t) \quad (14)$$

with $f_{i_k}(x_{i_k}; t)$ satisfying a one-dimensional linear equation

$$\partial_t f_{i_k}(x; t) = \left(\mathcal{L} + \gamma(t) - \frac{2n}{n+1} w \right) f_{i_k}(x; t) \quad (15)$$

subject to the initial conditions $f_{i_k}(x; 0) = c_{i_k}(x; 0)$.

Note that the decay of genetic diversity in the n th-tuned model is described by the spectrum of the operator appearing on the right-hand side of Equation 15. This operator also occurs as one of the terms in Equation 12, which describes

the fluctuations of the total population. Consistent with our population genetic interpretation, we can show quite generally that this operator has only decaying relaxation modes (see Figure B1), as Equation 15 misses the last positive source term of Equation 12.

We consider Equations 14 and 15 to be our formal results with the most immediate applicability because they provide a feasible approach to fundamental observables in population genetics that describe the maintenance and decay of genetic diversity, which *relies* on having access to higher moments.

After a few remarks and numerical tests of our formal results, we will thus focus on obtaining specific predictions of the decay of genetic diversity in models of adaptation and invasion.

Remarks

The above formalism includes the case $n = 1$ (closed first moment) presented in Hallatschek (2011) (in which the tuned weighting function $u^{(1)}$ was denoted by u_*).

In constrained random walk models, we can generally retrieve lower moments from higher moments by contraction with the weighting function u (in the variable x_m),

$$\begin{aligned} \langle u | C^{(n)} \rangle_m &= c(x_1; t) \dots \left(\int_{x_m} u(x_m; t) c_t(x_m; t) \right) \dots c(x_n; t) \\ &= c(x_1; t) \dots c(x_{m-1}; t) \, 1 \, c(x_{m+1}; t) \dots c(x_n; t) \\ &\equiv C^{(n-1)}(\backslash x_m; t), \end{aligned} \quad (16)$$

where we invoked the global constraint Equation 7 in going from the second to the third line. To simplify the notation, we have here introduced the shorthand $C^{(n-1)}(\backslash x_m; t)$ to denote $C^{(n-1)}(x_1, \dots, x_{m-1}, x_{m+1}, \dots, x_n; t)$, i.e., that the m th space variable should be omitted in the arguments.

The contraction rule Equation 16 thus shows that, if one obtains the n th moment by solving Equation 12, one generally has access to all lower-rank moments as well (but not to the larger moments). Thus, the model tuned to be linear at the n th moment gives access to all moments up to and including the n th.

It is remarkable that the governing equations for w and $\overline{C}^{(n)}$ are independent of the offspring number variation b . The only effect of increasing b is to reduce the fixation probability $u^{(n)} \propto b^{-1}$ and scale up the population sizes $\overline{C}^{(n)} \propto b^n$. This means in particular that noise-induced terms $\propto w$ in the moment equation are not scaled by b and thus do not become small in the small noise limit. The lack of a potentially small parameter has important consequences for the use of perturbation theory in this context: Number fluctuations are always nonnegligible in these dynamical systems.

Specific Results

We now illustrate our formal expressions by explicit numerical and analytical results for tuned models of traveling waves and compare them with fixed population size models. We thereby focus on models closed at the first and the second

Figure 2 Numerical solution of the model of adaptation tuned to be closed at the second moment ($n = 2$). (A and B) The mean population densities of the models with $n = 2$ (purple) and $n = 1$ (blue) for the same wave speed ($vD^{-1/3} = 6$) almost coincide. The profile for $n = 2$ (and generally for larger n) is slightly shifted, but the population size $\bar{N}$ is almost unaffected (see also Figure 4C), with even fewer deviations for larger population sizes (or adaptation speeds). Profiles from stochastic simulations subject to a fixed population size constraint are shown as brown circles. For most of the profile, the standard error is smaller than the circle size. A Gaussian approximation to the population density is delineated in gray (with $\bar{N}$ and v identical to the tuned parameters). Note from the semilogarithmic plot (B) that the tails of the tuned model and the fixed population size simulations deviate from the Gaussian profile in a similar fashion. (C) The numerical solution of the second moment $\bar{C}^{(2)}(x, y)$ ($n = 2$) is shown in a semilogarithmic plot. The ridge $\bar{C}^{(2)}(x = y)$ is nearly Gaussian (in logarithmic scale) except for its exponential tails. (D) From the second moment, we calculate the Pearson product-moment correlation $\rho(x, y)$ (see main text), which allows us to distinguish correlated (+1), uncorrelated (0), and anticorrelated (−1) variation. The nose of the wave clearly is anticorrelated with the bulk of the wave in our tuned models.

order to elucidate both the mean behavior of the traveling wave and spatiotemporal two-point correlations.

The two population biological phenomena we consider, adaptation and invasion, both give rise to compact traveling waves in fitness space and real space, respectively. They fundamentally differ in their position-dependent growth rates: While growth rates are linearly increasing toward the tip of the waves in adaptation models, they saturate in invasion models.

Models of adaptation

The detailed behavior of simple models of asexual adaptation varies with the assumptions made about the frequency of mutations and their fitness effects. In our numerical examples, we focus on the diffusion kernel in Equation 3, which assumes that a growth rate variance D is acquired per generation due to mutations. The advantage of the diffusion kernel is that it depends on only one parameter, the diffusion coefficient D , and matters when mutation rates are large compared to the (effective) rate of selection (Tsimring *et al.* 1996; Goyal *et al.* 2012).

Numerical approach: To solve our framework of tuned models, we used a multidimensional Newton–Raphson method to determine traveling-wave solutions corresponding to tuned models of degree n . To this end, we first determined a traveling steady-state solution $w(x; t) = w(x - vt)$ of Equation 11, defining the set of tuned weighting functions $u^{(n)}(x; t) = 2w(x; t)/b(n + 1)$ (Equation 10). Such a weighting function enforces also a traveling steady state for the n th moment $\bar{C}^{(n)}(x_1, \dots, x_n; t)$ of the number densities, which satisfies the linear time-independent Equation 12. Details of the numerical scheme and the explicit equations solved for $n = 1$ and $n = 2$ are described in Appendices C and D.

Model tuned to the second moment: Results for $n = 1$ have been presented and analyzed in detail in Hallatschek (2011). Figure 2 characterizes the behavior of $n = 2$ models in comparison with $n = 1$ tuned models. Since $n = 2$ models provide access to the second moment, one can of course also obtain the first moment by contraction with the weighting function $u^{(n)}(x, t)$ (cf. Equation 12). Figure 2A shows the mean population densities for both $n = 1$ and $n = 2$ models for the same velocity and weighting function w . The respective functions $u^{(1)}$ and $u^{(2)}$ are identical up to an n -dependent factor, cf. Equation 10. The mean population densities can hardly be distinguished in the figures—generally, the agreement of the two different models increases with increasing wave velocity or, equivalently, population size. Moreover, the stochastic simulations with a fixed population size track the predicted mean in good agreement, providing support for our claim that choice of a particular constraint becomes negligible for large population sizes.

Figure 2B shows the second moment $\bar{C}^{(2)}(x, y)$ in a semi-logarithmic three-dimensional plot. While the bell-shaped graph of the second moment is mostly Gaussian, it reveals a distinctly exponential decay in the front and back of the wave. This indicates the importance of fluctuations in these low-density regions.

More important than the absolute fluctuations are fluctuations relative to the mean. These are encoded, for instance, in the Pearson product-moment correlation, defined as $\rho(x, y) = \text{Cov}[c_t(x), c_t(y)] / \sqrt{\text{Var}[c_t(x)]\text{Var}[c_t(y)]}$. This quantity is depicted in Figure 2C and shows a clear anticorrelation signal in the nose of the wave. This signal arises from instances where, by chance, an individual catches a mutation that propels unusually far into nose of the wave. Then, one has an exponentially growing clone that accumulates a large fixation probability. To comply with the global constraint that the

Figure 3 Fixation in traveling-wave models of adaptation sensitively depends on individuals in the nose of the wave. One of our key results is that tuned models generally have the convenient feature that fixation probabilities can be computed exactly—they are given by the tuned weighting function $u^{(n)}(x; t) = 2w(x; t)/b(n+1)$. Here, we provide a test of this prediction. First, we have generated 50 independent start configurations and labeled the wave tip such that the labeled population is predicted to have a 50% fixation probability, $\langle u^{(1)} | c_t \rangle = 0.5$. (A) One start configuration and the labeled subpopulation in orange. [The shown density profile (light blue area) is momentarily smaller than the mean density (blue line) due to random population size fluctuations; cf. Figure 4]. For each of the labeled start configurations, we ran 200 simulations up until the (fixation) time where the labeled subpopulation either is extinct or has taken over. From counting the number of extinction and fixation events, Bayesian inference allows us to compute the posterior distribution, assuming a flat prior (Geyrhofer 2014). (B) For the model with $n = 1$, we find that 45 of the 95%-confidence intervals of this Beta-distributed posterior include the expected fixation probability of $1/2$ (black intervals), while 5 intervals scatter considerably due to lattice effects in simulations (red intervals). (C) When closing the moment hierarchy at $n = 2$, results exhibit stronger scattering. However, the measured confidence intervals for $n = 2$ reproduce the expected outcome to a reasonable extent.

total probability of fixation remains equal to 1, Equation 7, the entire population needs to be culled as the unusually fit clone in the front grows. The measured anticorrelation between nose and bulk of the wave thus reflects the fact that fixation probabilities are largest in the tip of the wave. These features and, hence, the anticorrelation are generic for constrained, pulled waves (Geyrhofer and Hallatschek 2015).

Fixation probabilities: The numerical solution for the weighting functions $u^{(n)}(x; t) = 2w(x; t)/b(n+1)$ in Figure 2 shows a strong increase toward the tip of the wave where they cross over to a linear increase. This behavior is consistent with the interpretation of a fixation probability: The success probability of an individual should be much larger in the tip of the wave rather than the bulk because it has to compete with less and less equally or more fit individuals. The fixation probability approaches a linear branch beyond some crossover in fitness because these individuals are so exceptionally fit that they merely have to survive random death to fix. The competition with conspecific is minimal—they merely have to compete with their own “relatives,” which descend from the same ancestor that mutated, at some point, far out into the nose of the wave.

To test our prediction that $u^{(n)}(x)$ can be interpreted, exactly, as a fixation probability, we have carried out the follow-

Figure 4 Fluctuations in adaption waves suggest the adaptation speed is controlled by the mean logarithmic population size. (A) Stochastic simulations reveal that, for a given speed of adaptation, tuned models have a larger mean population size $\bar{N}$ (blue line) than models of fixed population size (green line). The discrepancy between these two models is small compared to the best available cutoff theory [gray line (Cohen *et al.* 2005b)]. Most of the remaining discrepancy is due to population size fluctuations (very nearly exponentially distributed, see B): When we measure fluctuations in a stochastic realization of the tuned model and impose exactly the same population size time series in repeated stochastic simulations, we obtain an adaptation speed vs. a mean population size relation that coincides with the one obtained for the tuned models (red circles). Importantly, fixed population size and tuned models have almost the same mean logarithmic population size, $\log \bar{N}$, for the same wave speeds. It can be shown that this is due to a timescale separation between slowly decaying population size fluctuations (coalescence timescale) and the fast relaxation of the wave speed (determined by a mixing time in the tip of the wave). Thus, the key dynamical quantity in traveling-wave models with fluctuating population sizes is $\log N_t$ instead of $\bar{N}_t$. (C) The population size in models tuned for $n = 2$ varies only slightly compared to results with identical adaptation speeds obtained for $n = 1$. The ratio $\bar{N}^{(n=1)}/\bar{N}^{(n=2)}$ deviates only by 10% (at maximum) from 1 and converges to identical population sizes for increasing adaptation speeds. In addition, when using the constraint $u^{(2)}$ to control population sizes in stochastic simulations, we observe the identical exponential tail for $\mathbb{P}[N/\bar{N}]$ as for $n = 1$.

ing test: We ran simulations for $n = 1$ and $n = 2$ tuned models, in which we labeled the tip of the wave such that the predicted fixation probability is exactly 50% (cf. Figure 3). The measured fixation probability is shown in the insets of Figure 3. Within the statistical error, the agreement is very good.

We point out that the fact that $w(x; t)$ is strongly increasing toward the wave tip means that highly fit individuals have a large impact on what fraction of the population is culled per time step. For instance, if a single individual arises far out in the tail of the wave, it may have a large fixation probability. To keep the total fixation probability equal to 1, a significant fraction of the individuals needs to be cleared from the population. Importantly, however, the culling itself is independent of the identity of the individuals; *i.e.*, a poorly fit individual is equally likely to die as a highly fit individual. This distinguishes our constrained branching random walks from models that regularize the population size by removing cells preferentially at the wave tip (Derrida and Simon 2007; Berestycki *et al.* 2013). Such a procedure can modify the wave dynamics and, in particular, reduce the amount of fluctuations.

The speed of adaptation: Figure 4 shows the relation between velocity and mean population size in tuned models. For comparison, we also show the corresponding relationship

Figure 5 Timescales for the decay of genetic diversity in models of adaptation. In models with limited population size, the ultimate fate of a subpopulation of neutral mutants is to either go extinct or go to fixation. Hence, the diversity of the population will gradually decay unless new mutations come in, and quantifying this decay is an important population genetic challenge. A and B show as a function of velocity and population size, respectively, the leading decay times τ_0 and τ_1 of the function $f(x, t)$ in Equation 15, which controls the decay of genetic diversity via Equation 14. (A) Our numerical results suggest that, for the considered model with a diffusion kernel, the longest decay time approaches $\tau_0^{(n=2)} \propto v \sim (\ln N)^{1/3}$ for large speeds (gray line indicates the approximation in Equation B40 for $n = 2$). By contrast, the second longest decay time $\tau_1^{(n=2)}$ approaches a constant for large populations. This indicates an important timescale separation, as we argue in the main text. Also note that the numerical results approach, for large N , the approximation $\tau_{0, \text{approx}}$ in Equation B21, which is based on the timescale separation of the two lowest eigenvalues emerging for large population sizes. To corroborate these predictions we also conducted simulations tracking the so-called heterozygosity in stochastic simulations. Twice the measured decay time $\tau_{\text{heterozyg}}$ of the (asymptotic) exponential decay is plotted as blue circles. The factor of 2 originates from the fact that the measured heterozygosity is a two-point correlation function, which is quadratic in the function $f(x, t)$ as explained in Appendix B. The actual decay of the heterozygosity is shown in Figure B2C.

for fixed population size models, for which the velocity needs to be averaged. These models correspond to different statistical ensembles, yet for large population sizes the curves approach each other very well. Moreover, the influence of different n is small, as we show in an exemplary comparison of population sizes for $n = 1$ and $n = 2$ in Figure 4C.

The differences between the different curves in Figure 4A can be largely explained by population size fluctuations: If we force a fixed population size model to fluctuate precisely as a given realization N_t of a tuned model, we obtain very accurate agreement. Conversely, if we plot velocity vs. $\overline{\log N}$, we obtain improved agreement between all models. This agreement is due to a timescale separation, discussed below.

Decay of genetic diversity: In general, correlation functions $\overline{C}^{(n)}$ describe fluctuations of the population as a whole, and these affect the genetic diversity of the population. The decay of diversity can be measured in experiments, for instance, by tracking the dynamics of distinguishable subpopulations, identifiable via genetic or fluorescent markers. Equations 14 and 15 describe the cross-correlation function between n such differently labeled subpopulations. The cross-correlation function for two subpopulations is closely related to the so-called heterozygosity in population genetics, as detailed in Appendix B. We

have solved Equation 15 numerically to determine how fast the heterozygosity decays in our model of adaptation.

Figure 5 depicts the behavior of the largest two relaxation times for two subpopulations ($n = 2$) as a function of the velocity of the wave and the population size, respectively. See also Figure B1 for a more detailed relaxation spectrum.

The asymptotic behavior of the decay of genetic diversity can be found analytically, as we show in Appendix B. We find that genetic diversity decays on the timescale $\tau_0 \sim v / (2D \ln(4/3))$ (for $n = 2$). Our approximation, including the coefficient, agrees very well with the numerical results in Figure 5. (A comparison with stochastic simulations is shown in Figure B2, B and C.) For very large speeds, our prediction approaches $\tau_0 \sim (\ln N)^{1/3}$ in terms of the population size. This rigorous scaling result is consistent with the measured time constant for coalescence in our model (Neher and Hallatschek 2013), which has been previously explained on the bases of heuristic arguments.

Interestingly, we find that while the longest relaxation time τ_0 increases linearly with wave speed, the next higher relaxation time approaches a constant value, $\tau_1 \sim D^{1/3}$. The timescale separation $\tau_0 \gg \tau_1$ for $vD^{-2/3} \gg 1$ can be used to derive a very general approximation for the timescale for the decay of genetic diversity (Appendix B), $\tau_0 \sim \langle w^2 | \bar{c} \rangle^{-1}$, which is also shown in Figure 5. Moreover, we hypothesize that this timescale separation could be relevant for the question of why the Bolthausen–Sznitman coalescent is found in many models of adaptation and invasion waves of the Fisher–Kolmogorov type (Brunet *et al.* 2007; Desai *et al.* 2013; Neher and Hallatschek 2013).

Details of the relaxation spectrum of our model, including the form of the eigenfunctions and the timescale separation, can be understood from an analogy to the physical problem of a (quantum) particle trapped in a potential well, as we describe in Appendix B.

Invasion waves

After changing the Liouville operator $\mathcal{L}$ to the one in Equation 8, and following the same numerical procedures as described above (Appendix C), we obtain analogous results for invasion waves (see Figure 6). Note that the spatial coordinate now corresponds to real space rather than fitness space. Our data reproduce the universal velocity–population size relationships that have been established for FKPP waves over the last 20 years (van Saarloos 2003; Brunet *et al.* 2006).

Our numerics show that the leading decay rate of the heterozygosity follows the scaling $\tau_0 \sim s^{-1} [\ln N(D/s)^{1/2}]^3$, which is reproduced by our separation-of-timescale approximation. This result is consistent with the time constant for coalescence in the wave tip obtained by a phenomenological approach (Brunet *et al.* 2007).

Discussion

The ecological and evolutionary fate of populations often depends on a small number of pioneers, distinguished by their

Figure 6 Numerical results for invasion waves. (A) Wave speed as a function of average population size in the model tuned for the first moment, $n = 1$. The gray line compares our results with the leading-order correction $2 - v\sqrt{Dr} \approx \pi^2/(\ln N)^2$ derived from phenomenological arguments (Brunet and Derrida 1997; Brunet *et al.* 2006). Inset B indicates the approach of numerical data to this correction for even larger population sizes on a double logarithmic axis. C is an analog to Figure 5 and shows the two longest decay times, $\tau_0 > \tau_1$, of the genetic diversity, as measured by two-point correlation functions in

tuned models (with $n = 2$). The increasing gap between both decay times manifests a timescale separation. The double logarithmic plot in inset D shows that the longest decay time follows $\tau_0 \sim (\ln N)^3$ (upper gray lines) asymptotically and, hence, behaves as the predicted coalescence time in FKPP waves and the closely related branching random walks with absorption (Derrida and Simon 2007; Berestycki *et al.* 2013). By contrast, the second decay time follows $\tau_1 \sim (\ln N)^2$ (lower gray line).

growth rates, their location, or other characteristics correlated with long-term survival. Most analyses of these inherently stochastic problems have focused on the mean behavior of the population. Yet the mean behavior says little about population-level fluctuations in these models, which control, for instance, the decay of the genetic diversity in the population.

Here, we have shown that fluctuations can be analyzed in principle, if one relies on minimal models that reduce the dynamics to two essential ingredients: (1) Birth, death, and jumps give rise, effectively, to a branching random walk and (2) a nonlinear population regulation makes sure that those branching processes do not get out of control. For such constrained branching random walks, we have provided a general route toward analyzing fluctuations. The basic idea of the method is to adjust the population control, an essential non-linearity, in such a way that the equations describing correlation functions of order n are closed and linear.

We have used our method to compute the decay of genetic diversity induced by random genetic drift. To obtain explicit results, we have focused on simple models of adaptation and of species invasions. We found that genetic diversity, as measured by the heterozygosity, decays exponentially on long times. The decay time increases only very slowly with population size, as a power of the logarithm of the population size. This is in marked contrast to neutral Wright–Fisher populations where the decay of genetic diversity is expected to last of order N generations.

In the case of adaptation waves with a diffusive mutation kernel, we could derive the particularly simple result that heterozygosity decays on the timescale $\tau_0 \sim v/D$, if many mutations of weak effect drive the adaptation. This relation could be useful for the analysis of evolution experiments: The decay of genetic diversity can be measured in evolution experiments, for instance, by tracking to two differently labeled subpopulations in replicate experiments. If one then regresses τ_0 on the speed v of adaptation, one could infer the fitness variance D per generation that is generated by mutations. Genetic barcoding techniques allow tracking of many subpopulations (Levy *et al.* 2015). This will make it possible to study, in a similar way, the higher-order cross-correlation functions in Equation 13, which are generalizations of the heterozygosity.

As another notable fluctuation effect, we have detected a marked anticorrelation between the dynamics in the tip and the bulk of the wave. This peculiar signal reflects the fact that there can be only one fixing clone and that fixing probabilities are exponentially large in the tip of the wave. We expect these features and the anticorrelation to be generic for constrained, pulled waves (Geyerhofer and Hallatschek 2015). However, we anticipate the details of these anticorrelations to be model dependent, for instance, when fluctuations in the tip region immediately lead to fluctuations of the total population size when the constraint has a lot of weight in the tip of the wave, as it has in our tuned models. For fixed population size models, on the other hand, fluctuations in the tip need some time to influence the total population size. Thus, we expect anticorrelations require a certain delay time to build up in fixed population size models. The delay time will be related to the time it takes for the traveling wave to move by the distance between the bulk region of the wave with maximal population density and the tip of the wave, where the pioneers reside (Neher and Hallatschek 2013).

Moreover, we found that, for the models analyzed, the timescale for the decay of higher-order correlations, such as the genetic heterozygosity, is substantially longer than the timescales for wave shape relaxation. The presence of such a timescale separation simplifies the fluctuation analysis considerably and may underlie the observation that various pulled waves are characterized by similar multiple-merger coalescence processes (Brunet *et al.* 2006; Desai *et al.* 2013; Neher and Hallatschek 2013).

One might wonder about the net effect of noise on models of adaptation and other traveling waves. If one is concerned only with the mean, many previous works have assumed that the effect of noise can be summarized by an effective cutoff in the tip of the wave (Tsimring *et al.* 1996; Brunet and Derrida 1997; Cohen *et al.* 2005b). This cutoff effect can be explicitly seen in the closed first moment equation of tuned models, as was already pointed out in Hallatschek (2011). However, what is the effect of noise on higher-order correlations? Our general formulation in Equation 12 of the n th moment exhibits, in general, two terms with different signs that are unexpected in a deterministic framework. One term tends to generate positive correlations between nearby individuals (in

fitness space). These correlations then dissipate over distances due to the term with negative sign. Importantly, the correlation term dominates in the tip of the wave due to its density dependence. The net effect of fluctuations on the correlations emphasizes the complex nature of fluctuations, which only to the lowest order can be captured by a simple cutoff term in an effective Liouville operator.

The branching random walk contains a parameter b , the variance in offspring numbers per generation, that effectively measures the strength of genetic drift. The larger b is, the lower the establishment probability of beneficial mutations. Surprisingly, the noise-induced terms in our moment equations do not depend on this parameter b . This means that the noise-induced effects are not small, in general, even if the parameter b is small, such that a controlled small-noise perturbation analysis is not possible. This reinforces the observation that noise is a singular perturbation that fundamentally affects the outcome of ecological and evolutionary processes.

In principle, our method of model tuning applies to any branching random walk subject to a global constraint. This includes models that combine ecology and evolution (Pelletier *et al.* 2009; Barton *et al.* 2013) or epistatic models in which mutations are not simply additive but might interact (Tenaillon 2014). However, for more complex scenarios of interest to evolutionary biologists, we want to introduce additional nonlinearities. For instance, sex or recombination introduces a quadratic nonlinearity as it depends on the probability density of two different individuals finding each other and mating (Neher *et al.* 2010). In evolutionary game theory, we are interested in mutants that have a frequency-dependent advantage (Kussell and Vucelja 2014; Reiter *et al.* 2014). The fitness of producers of a common good depends on the frequency of producers. This, again, introduces nonlinearities, which are quadratic in the simplest case.

Such nonlinearities cannot be included in an exact way because they generate higher-order terms. However, it may be a useful approach to build them in and truncate the moment hierarchy at an appropriate order provided one can show that the neglected terms really are small. Such an approach could work if the nonlinearities do not strongly influence the dynamics in the small density regions where the noise strength is large. A truncation scheme would, in this case, amount to matching a stochastic but linear description of the wave tip with a deterministic but nonlinear bulk of the wave. These difficult but crucial nonlinear problems may be fruitful topics for future work.

Acknowledgments

Thanks to Peter Pfaffelhuber for making us aware of Katzenberger (1991). This work was partially supported by a Simons Investigator award from the Simons Foundation (OH), and the Deutsche Forschungsgemeinschaft via grant HA 5163/2-1 (OH).

Literature Cited

- Allen, L. J., 2008 *An Introduction to Stochastic Processes with Applications to Biology*. CRC Press, Boca Eaton, FL.
- Barroso-Batista, J., A. Sousa, M. Lourenço, M.-L. Bergman, D. Sobral *et al.*, 2014 The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet.* 10: e1004182.
- Barton, N., 1979 The dynamics of hybrid zones. *Heredity* 43: 341–359.
- Barton, N., A. Etheridge, and A. Véber, 2013 Modelling evolution in a spatial continuum. *J. Stat. Mech.* 2013: P01002.
- Berestycki, J., N. Berestycki, and J. Schweinsberg, 2013 The genealogy of branching Brownian motion with absorption. *Ann. Probab.* 41: 527–618.
- Brunet, E., and B. Derrida, 1997 Shift in the velocity of a front due to a cutoff. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 56: 2597.
- Brunet, E., B. Derrida, A. Mueller, and S. Munier, 2006 Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73: 056126.
- Brunet, E., B. Derrida, A. Mueller, and S. Munier, 2007 Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76: 041104.
- Cohen, E., D. A. Kessler, and H. Levine, 2005a Fluctuation-regularized front propagation dynamics in reaction-diffusion systems. *Phys. Rev. Lett.* 94: 158302.
- Cohen, E., D. A. Kessler, and H. Levine, 2005b Front propagation up a reaction rate gradient. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72: 066126.
- Derrida, B., and D. Simon, 2007 The survival probability of a branching random walk in presence of an absorbing wall. *Europhys. Lett.* 78: 60006.
- Desai, M. M., and D. S. Fisher, 2007 Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics* 176: 1759–1798.
- Desai, M. M., D. S. Fisher, and A. W. Murray, 2007 The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* 17: 385–394.
- Desai, M. M., A. M. Walczak, and D. S. Fisher, 2013 Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* 193: 565–585.
- Fisher, D. S., 2013 Asexual evolution waves: fluctuations and universality. *J. Stat. Mech.* 2013: P01011.
- Fleming, W. H., 1979 Equilibrium distributions of continuous polygenic traits. *SIAM J. Appl. Math.* 36: 148–168.
- Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66: 95–107.
- Gerrish, P. J., and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* 102: 127–144.
- Geyrhofer, L., 2014 Quantifying evolutionary dynamics. Ph.D. Thesis, Max Planck Institute for Dynamics and Self-Organization and University of Göttingen, Göttingen, Germany.
- Geyrhofer, L., and O. Hallatschek, 2013 Stochastic delocalization of finite populations. *J. Stat. Mech.* 2013: P01007.
- Good, B. H., and M. M. Desai, 2013 Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theor. Popul. Biol.* 85: 86–102.
- Good, B. H., I. M. Rouzine, D. J. Balick, O. Hallatschek, and M. M. Desai, 2012 Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci. USA* 109: 4950–4955.
- Gordo, I., L. Perfeito, and A. Sousa, 2011 Fitness effects of mutations in bacteria. *J. Mol. Microbiol. Biotechnol.* 21: 20–35.

- Goyal, S., D. J. Balick, E. R. Jerison, R. A. Neher, B. I. Shraiman *et al.*, 2012 Dynamic mutation–selection balance as an evolutionary attractor. *Genetics* 191: 1309–1319.
- Haccou, P., P. Jagers, and V. A. Vatutin, 2005 *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge University Press, Cambridge/London/New York.
- Hallatschek, O., 2011 The noisy edge of traveling waves. *Proc. Natl. Acad. Sci. USA* 108: 1783–1787.
- Hallatschek, O., and K. Korolev, 2009 Fisher waves in the strong noise limit. *Phys. Rev. Lett.* 103: 108103.
- Hallatschek, O., and D. R. Nelson, 2008 Gene surfing in expanding populations. *Theor. Popul. Biol.* 73: 158–170.
- Hallatschek, O., P. Hersen, S. Ramanathan, and D. R. Nelson, 2007 Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. USA* 104: 19926–19930.
- Hedgcock, D., and A. I. Pudovkin, 2011 Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull. Mar. Sci.* 87: 971–1002.
- Hegreness, M., N. Shores, D. Hartl, and R. Kishony, 2006 An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* 311: 1615–1617.
- Johnson, T., and N. Barton, 2005 Theoretical models of selection and mutation on quantitative traits. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1411–1425.
- Katzenberger, G. S., 1991 Solutions of a stochastic differential equation forced onto a manifold by a large drift. *Ann. Probab.* 19: 1587–1628.
- Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri *et al.*, 2012 Experimental evolution. *Trends Ecol. Evol.* 27: 547–560.
- Kimura, M., 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Natl. Acad. Sci. USA* 54: 731.
- Kussell, E., and M. Vucelja, 2014 Non-equilibrium physics and evolution—adaptation, extinction, and ecology: a key issues review. *Rep. Prog. Phys.* 77: 102602.
- Lande, R., 1975 The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* 26: 221–235.
- Levy, S. F., J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher *et al.*, 2015 Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519: 181–186.
- Müller, M. J., B. I. Neugeboren, D. R. Nelson, and A. W. Murray, 2014 Genetic drift opposes mutualism during spatial population expansion. *Proc. Natl. Acad. Sci. USA* 111: 1037–1042.
- Nagylaki, T., 1978 Random genetic drift in a cline. *Proc. Natl. Acad. Sci. USA* 75: 423–426.
- Neher, R. A., 2013 Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu. Rev. Ecol. Evol. Syst.* 44: 195–215.
- Neher, R. A., and O. Hallatschek, 2013 Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci. USA* 110: 437–442.
- Neher, R. A., and B. I. Shraiman, 2012 Fluctuations of fitness distributions and the rate of Muller’s ratchet. *Genetics* 191: 1283–1293.
- Neher, R. A., B. I. Shraiman, and D. S. Fisher, 2010 Rate of adaptation in large sexual populations. *Genetics* 184: 467–481.
- Neher, R. A., T. A. Kessinger, and B. I. Shraiman, 2013 Coalescence and genetic diversity in sexual populations under selection. *Proc. Natl. Acad. Sci. USA* 110: 15836–15841.
- O’Hagan, A., and J. Forster, 1994 *Kendall’s Advanced Theory of Statistics: Bayesian Inference*, Vol. 2B. Halsted Press, New York.
- Pelletier, F., D. Garant, and A. P. Hendry, 2009 Eco-evolutionary dynamics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364: 1483–1489.
- Perfeito, L., L. Fernandes, C. Mota, and I. Gordo, 2007 Adaptive mutations in bacteria: high rate and small effects. *Science* 317: 813–815.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007 *The Art of Scientific Computing* (Numerical Recipes, Ed. 3). Cambridge University Press, Cambridge/London/New York.
- Reiter, M., S. Rulands, and E. Frey, 2014 Range expansion of heterogeneous populations. *Phys. Rev. Lett.* 112: 148103.
- Rouzine, I. M., J. Wakeley, and J. M. Coffin, 2003 The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci. USA* 100: 587–592.
- Rouzine, I. M., É. Brunet, and C. O. Wilke, 2008 The traveling-wave approach to asexual evolution: Muller’s ratchet and speed of adaptation. *Theor. Popul. Biol.* 73: 24–46.
- Schiffels, S., G. J. Szöllösi, V. Mustonen, and M. Lässig, 2011 Emergent neutrality in adaptive asexual evolution. *Genetics* 189: 1361–1375.
- Slatkin, M., 1985 Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* 16: 393–430.
- Tenaillon, O., 2014 The utility of Fisher’s geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 45: 179–201.
- Tsimring, L. S., H. Levine, and D. A. Kessler, 1996 RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* 76: 4440.
- Turelli, M., 1984 Heritable genetic variation via mutation–selection balance: Lerch’s zeta meets the abdominal bristle. *Theor. Popul. Biol.* 25: 138–193.
- Van Kampen, N. G., 1992 *Stochastic Processes in Physics and Chemistry*, Vol. 1. Elsevier, Amsterdam/New York.
- van Saarloos, W., 2003 Front propagation into unstable states. *Phys. Rep.* 386: 29–222.
- Wakeley, J., 2009 *Coalescent Theory: An Introduction*, Vol. 1. Roberts & Co., Greenwood Village, CO.
- Weber, M. F., G. Poxleitner, E. Heibisch, E. Frey, and M. Opitz, 2014 Chemical warfare and survival strategies in bacterial range expansions. *J. R. Soc. Interface* 11: 20140172.
- Weissman, D. B., and O. Hallatschek, 2014 The rate of adaptation in large sexual populations with linear chromosomes. *Genetics* 196: 1167–1183.

Communicating editor: N. H. Barton

Appendices

Appendix A: Basic Derivations

In the following we proceed with the derivation of our main results quoted in *Summary of Main Formal Results*. To this end, we first recapitulate the derivation of the stochastic differential equation governing the dynamics of branching random walks, as was done in Hallatschek (2011). We then discuss the consequences of these stochastic dynamics for n -point correlation functions, which will allow us to identify the natural choice of the weighting function $u^{(n)}$. Subsequently, we discuss how to modify our basic model to account for different subtypes within the population.

Stochastic dynamics of constrained branching random walks

We now determine the stochastic dynamics obtained in the limit $\epsilon \rightarrow 0$, which is in general nonlinear and hence not solvable. We then use the ensuing stochastic differential equation to identify special models with closed moment hierarchies. We will find that these tuned models not only are solvable, but also allow for a natural interpretation of the constraint in terms of fixation probabilities.

The stochastic dynamics of a constrained branching random walk (CBRW) were derived in Hallatschek (2011) and may be summarized as follows. The state of the system is described by the number density $c_t(x)$ of random walkers at position x and time t . At any time, the distribution of random walkers has to satisfy a global constraint defined by Equation 7.

The combination of Equations 1 and 7 can be written as a fraction,

$$c_{t+\epsilon} = \frac{c_t + \epsilon \mathcal{L}c_t + \sqrt{\epsilon b c_t} \eta_t}{\langle u_{t+\epsilon} | c_t + \epsilon \mathcal{L}c_t + \sqrt{\epsilon b c_t} \eta_t \rangle}, \quad (\text{A1})$$

in the continuous-time limit (small enough ϵ is required to ensure that the denominator of the fraction is never far from 1). Note that the expression in Equation A1 evidently satisfies the constraint, $\langle u_{t+\epsilon} | c_{t+\epsilon} \rangle = 1$.

In the eventual continuous-time limit, we need only to retain terms up to order $O(\epsilon)$ (Van Kampen 1992). Thus, expanding Equation A1, we obtain

$$\begin{aligned} c_{t+\epsilon} = c_t &+ \sqrt{\epsilon} \left[\sqrt{b c_t} \eta_t - c_t \langle u_t | \sqrt{b c_t} \eta_t \rangle \right] \\ &+ \epsilon \left[\mathcal{L}c_t - \sqrt{b c_t} \eta_t \langle u_t | \sqrt{b c_t} \eta_t \rangle \right] \\ &- \epsilon \left[c_t \langle \partial_t u_t | c_t \rangle + c_t \langle u_t | \mathcal{L}c_t \rangle - c_t \langle u_t | \sqrt{b c_t} \eta_t \rangle^2 \right]. \end{aligned} \quad (\text{A2})$$

In Equation A2, we required u_t to change only deterministically, $u_{t+\epsilon} = u_t + \epsilon \partial_t u_t + o(\epsilon)$; i.e., it has no stochastic component.

Finally, upon replacing products of order $\eta_t \eta_t$ in the deterministic term of expansion, Equation A2, with their averages, Equation 4, we obtain the following stochastic differential equation: The temporal change $\Delta c_t \equiv c_{t+\epsilon} - c_t$ of the concentration field c_t from time t to $t + \epsilon$ can be written as

$$\Delta c_t(x) \equiv c_{t+\epsilon} - c_t = \epsilon \Delta^{(d)} c_t(x) + \sqrt{\epsilon} \Delta^{(s)} c_t(x), \quad (\text{A3})$$

which consists of a deterministic change $\epsilon \Delta^{(d)} c_t$ of order $O(\epsilon)$ and a stochastic change $\sqrt{\epsilon} \Delta^{(s)} c_t$ of order $O(\epsilon^{1/2})$. These are given by

$$\begin{aligned} \Delta^{(d)} c_t(x) &= (\mathcal{L}_t - b u_t) c_t - \left\langle (\partial_t + \mathcal{L}_t^\dagger - b u_t) u_t \middle| c_t \right\rangle c_t, \\ \Delta^{(s)} c_t(x) &= \eta_t \sqrt{b c_t} - \langle u_t | \eta_t \sqrt{b c_t} \rangle c_t. \end{aligned} \quad (\text{A4})$$

Thus, we have arrived at the continuous-time stochastic process for a constrained branching random walk (Hallatschek 2011), which summarizes the combined effect of the original two-step algorithm, (i) “branching random walk” and (ii) “enforce constraint.” The related concept of forcing the solution of a stochastic differential equation onto a manifold has been analyzed in Katzenberger (1991).

Moment equations

This section introduces the hierarchy of moment equations of CBRWs that characterize the mean and the fluctuations of the concentration field of the random walkers.

Consider the dynamics of the products of $C_t^{(n)}$. Our goal is to determine how the function $C_t^{(n)}$ changes as time progresses. Since $C_t^{(n)}$ is the product of n density fields for which we have a stochastic differential equation, we could obtain the dynamics $C_t^{(n)}$ by application of Ito’s formula for nonlinear variable substitutions in stochastic differential equations. Equivalently, we can express the time increments of the n -point products $C_t^{(n)}$ in terms of the changes of the single fields c_t ,

$$\begin{aligned}
C_{t+\epsilon}^{(n)} &= \prod_{j=1}^n c_{t+\epsilon}(x_j) \\
&= \prod_{j=1}^n \left[c_t(x_j) + \epsilon \Delta^{(d)} c_t(x_j) + \sqrt{\epsilon} \Delta^{(s)} c_t(x_j) \right],
\end{aligned} \tag{A5}$$

using the deterministic and stochastic time increments $\Delta^{(d/s)} c_t$ computed in Equation A4. Next, we expand the product up to order $O(\epsilon)$,

$$\begin{aligned}
\Delta C_t^{(n)} &= \sqrt{\epsilon} \sum_{j=1}^n \sum_{j=1}^n C_t^{(n-1)}(\setminus x_j) \Delta^{(s)} c_t(x_j) \\
&\quad + \epsilon \left[\sum_{j=1}^n C_t^{(n-1)}(\setminus x_j) \Delta^{(d)} c_t(x_j) + \sum_{j=1}^{n-1} \sum_{k=j+1}^n C_t^{(n-2)}(\setminus x_j, \setminus x_k) \Delta^{(s)} c_t(x_j) \Delta^{(s)} c_t(x_k) \right].
\end{aligned} \tag{A6}$$

To simplify the notation, we use the shorthand $C^{(n-1)}(\setminus x_m; t)$ to denote $C^{(n-1)}(x_1, \dots, x_{m-1}, x_{m+1}, \dots, x_n; t)$, i.e., that the m th space variable should be omitted in the arguments.

Note that the last term arises only for $n \geq 2$. Inserting the time increments in Equation A4 of the single fields into Equation A6 yields

$$\begin{aligned}
\Delta C_t^{(n)} &= \sqrt{\epsilon} \left[\sum_{j=1}^n \sqrt{bc_t(x_j)} \eta_t(x_j) C_t^{(n-1)}(\setminus x_j) - n \langle u_t | \eta_t \sqrt{bc_t} \rangle C_t^{(n)} \right] \\
&\quad + \epsilon \left[\sum_{j=1}^n (\mathcal{L} - bu_t)|_{x_j} C_t^{(n)} - n \langle (\partial_t + \mathcal{L}^\dagger - bu_t) u_t | c_t \rangle C_t^{(n)} \right] \\
&\quad + \epsilon \sum_{j=1}^{n-1} \sum_{k=j+1}^n b \sqrt{c_t(x_j) c_t(x_k)} \eta_t(x_j) \eta_t(x_k) C_t^{(n-2)}(\setminus x_j, \setminus x_k) \\
&\quad - \epsilon(n-1) \sum_{j=1}^n \sqrt{bc_t(x_j)} \eta_t(x_j) \langle u_t | \eta_t \sqrt{bc_t} \rangle C_t^{(n-1)}(\setminus x_j) \\
&\quad + \epsilon \frac{n(n-1)}{2} \langle u_t | \eta_t \sqrt{bc_t} \rangle^2 C_t^{(n)}.
\end{aligned} \tag{A7}$$

Within the deterministic $O(\epsilon)$ terms, we again replace products of noise in terms of their averages, as given by Equation 4. Note that we can always write $\eta_t(x) \eta_t(y) = \delta_{tt'} \delta(x-y)$ plus a stochastic component. If such terms quadratic in the noise arise in the deterministic $O(\epsilon)$ part of any stochastic equation, one may simply ignore their stochastic component, as it would lead to (in the limit $\epsilon \rightarrow 0$) negligible contributions (Van Kampen 1992). We then obtain

$$\begin{aligned}
\Delta C_t^{(n)} &= \sqrt{\epsilon} \left[\sum_{j=1}^n C_t^{(n-1)}(\setminus x_j) \sqrt{bc_t(x_j)} \eta_t(x_j) - n C_t^{(n)} \langle u | \eta_t \sqrt{bc_t} \rangle \right] \\
&\quad + \epsilon \left[\sum_{j=1}^n (\mathcal{L} - bnu_t)|_{x_j} C_t^{(n)} + b \sum_{j=1}^n \sum_{k=j+1}^n \delta(x_j - x_k) C_t^{(n-1)}(\setminus x_k) \right] \\
&\quad - \epsilon n \langle (\partial_t + \mathcal{L}^\dagger - b(n+1)/2 u_t) u_t | c_t \rangle C_t^{(n)}.
\end{aligned} \tag{A8}$$

Upon averaging and sending ϵ to zero, we obtain an equation of moment for the n th moment,

$$\begin{aligned}
\partial_t \bar{C}_t^{(n)} &= \sum_{j=1}^n (\mathcal{L} - bnu_t)|_{x_j} \bar{C}_t^{(n)} + b \sum_{j=1}^n \sum_{k=j+1}^n \delta(x_j - x_k) \bar{C}_t^{(n-1)}(\setminus x_k) \\
&\quad - n \langle (\partial_t + \mathcal{L}^\dagger - b(n+1)/2 u_t) u_t | \bar{C}_t^{(n+1)} \rangle_{n+1}.
\end{aligned} \tag{A9}$$

From (A9) we observe that the coupling to higher moments $\bar{C}_t^{(n+1)}$ is mediated via the dynamics of the weighting function u_t in the last term.

Exact closure

The key result of Hallatschek (2011) was that a particular choice of the weighting function u_t exists, for which the first moment equation is closed. We now show that exact closures can be found for higher moments as well.

Suppose the solution $u_t^{(n)}(x)$ of

$$-\partial_t u_t^{(n)} = \left[\mathcal{L}^\dagger - \frac{b(n+1)}{2} u_t^{(n)} \right] u_t^{(n)} \quad (\text{A10})$$

exists, and we choose $u_t^{(n)}$ as the weighting function. For this particular model, the dependence on the $(n+1)$ th moment in Equation A9 disappears identically:

$$\partial_t \bar{C}_t^{(n)} = \sum_{j=1}^n \left(\mathcal{L} - b n u_t^{(n)} \right) \Big|_{x_j} \bar{C}_t^{(n)} + b \sum_{j=1}^n \sum_{k=j+1}^n \delta(x_j - x_k) \bar{C}_t^{(n-1)}(x_k). \quad (\text{A11})$$

Thus, the hierarchy of the first n moments is closed. In fact, this closed set of n differential equations can be summarized by a single integro-differential equation (Equation 12), because contracting $\bar{C}_t^{(n)}$ with $u_t^{(n)}$ reduces the order of the moments by virtue of the constraint (cf. Equation 16).

The final form of our results (Equations 11 and 12) is obtained upon substituting

$$u_t^{(n)}(x) = \frac{2w(x; t)}{b(n+1)}, \quad (\text{A12})$$

which is the initially quoted Equation 10.

In summary, starting from a linear operator $\mathcal{L}$, we have identified an algorithm to construct a constrained branching walk model solvable up to the n th moment: First, identify the weighting function $w(x; t)$ for the given operator $\mathcal{L}$. To this end, solve Equation A10, which is a deterministic nonlinear equation for the weighting function $w(x; t)$ depending on a space and a time variable. For the specific applications discussed in this article, the explicit time dependence is transformed away by the use of a comoving frame, in which correlation functions depend on the parameter combination $x - vt$ rather than on x and t separately. Second, choose n and solve the corresponding moment Equation 12, which is a linear equation for the function $\bar{C}^{(n)}$ that depends on n space variables and a time variable. Once the function $\bar{C}^{(n)}$ has been obtained, any lower-order moment follows by contraction with $u^{(n)} \propto w$, as described in Equation 16.

Appendix B: Accounting for Different Subtypes

We now extend our model to account for k different types of individuals. This enables studying questions such as, How does a mutator strain take over in an evolving population of bacteria even if it confers a direct fitness detriment? Or how does a faster dispersing mutant spread during a growing tumor even if it might be slower growing? Moreover, we can discuss the decay of genetic diversity and arrive at the very important conclusion that $w(x; t)$ is always the fixation probability of a neutral mutation arising at position x and time t by considering exchangeable subtypes.

To this end, we define the dynamics of the subtypes analogously to our original constrained branching walk model for the total population in Appendix A: The number density of individuals of type i at position x at time t is given by a density field $c_i(x, t)$. Hence, the entire state of the system is described by the vector $\vec{c}(x, t) = \{c_i(x, t)\}_{i \leq k}$. In each time step, a given subtype undergoes a step of branching random walk subject to its own linear dynamics encoded by an operator $\mathcal{L}_i$ and its own fluctuations generated by a noise field $\eta_i(x, t)$,

$$\tilde{c}_i(t + \epsilon) - c_i(t) = \epsilon \mathcal{L}_i c_i + \sqrt{\epsilon b_i c_i} \eta_i. \quad (\text{B1})$$

The i dependence of the linear operators $\mathcal{L}_i$ encodes the *phenotypic* differences between types. *E.g.*, if type i refers to a mutator type, $\mathcal{L}_i$ would include a particular mutational operator $\mathcal{M}$ characterizing the mutator phenotype. For instance, if mutations are modeled by diffusion, the corresponding diffusion constant of a mutator would be larger than that of the wild type.

The different subtypes are coupled only by the second computational substep

$$c_i(x, t + \epsilon) = \tilde{c}_i(x, t + \epsilon)(1 - \lambda), \quad (\text{B2})$$

which ensures a global constraint defined by a weighting function vector $\vec{u} = \{u_i\}_{i \leq k}$,

$$1 = \int dx \sum_{i=1}^k u_i(x, t) c_i(x, t) \equiv \langle \vec{u} | \vec{c} \rangle. \quad (\text{B3})$$

Note that we have merely added another (discrete) dimension to the problem—the type of degrees of freedom. It may be checked that our arguments to arrive at an effective stochastic differential equation and for closing the moment hierarchy in *Appendix A* generalize to any number of dimensions. Thus, we can immediately restate our central results for the extended model accounting for subtypes.

In particular, if we choose the weighting function vector to be $u_i^{(n)} = 2w_i/b_i(n+1)$ with

$$-\partial_t w_i(x, t) = \left[\mathcal{L}_i^\dagger + \gamma(t) - w_i(x, t) \right] w_i(x, t), \quad (\text{B4})$$

then the equation of motion for the n th moment will be closed. If we choose the notation

$$C_{i_1, \dots, i_n}^{(n)}(x_1, \dots, x_n; t) \equiv \prod_{j=1}^n c_{i_j}(x_j, t), \quad (\text{B5})$$

with i_j being the type of the j th number density field in the product on the right-hand side, the equation of motion for the n th moment reads

$$\partial_t \overline{C}_{i_1, \dots, i_n}^{(n)} = \sum_{j=1}^n \left(\mathcal{L}_{i_j} + \gamma(t) - \frac{2n}{n+1} w_{i_j} \right) \Big|_{x_j} \overline{C}_{i_1, \dots, i_n}^{(n)} + \frac{2}{n+1} \sum_{j=1}^n \sum_{l=j+1}^n \delta_{i_j, i_l} \delta(x_j - x_l) \langle \vec{w} | \overline{C}_{i_1, \dots, i_n}^{(n)} \rangle_l. \quad (\text{B6})$$

Note that the correlations indicated by the last term arise only for a subpopulation with itself, $i_j = i_l$.

The neutral case and interpretation of the weighting function

Now let us focus on the special case where types follow the same dynamics in the statistical sense,

$$\mathcal{L}_i = \mathcal{L}, \quad w_i(x; t) = w(x; t), \quad b_i = b. \quad (\text{B7})$$

For such “exchangeable” subtypes, it is easy to see that the equation of motion (Equation 12) for the n th moment of the total population, $c(x, t) = \sum_i c_i(x, t)$, is obtained upon summing the left- and right-hand sides of Equation B6 over all type indexes, i.e., by carrying out $\sum_{i_1} \sum_{i_2} \dots \sum_{i_n}$.

We can single out one particular subpopulation, say $i_1 = \ell$ (ℓ for labeled), by summing the equation of motion over all other type indexes, $\sum_{i_2} \dots \sum_{i_n}$. This yields

$$\partial_t \overline{c_\ell C^{(n-1)}} = \sum_{j=1}^n \left(\mathcal{L} + \gamma(t) - \frac{2n}{n+1} w \right) \Big|_{x_j} \overline{c_\ell C^{(n-1)}} + \frac{2}{n+1} \sum_{j=1}^n \sum_{k=j+1}^n \delta(x_j - x_k) \langle w | \overline{c_\ell C^{(n-1)}} \rangle_k. \quad (\text{B8})$$

Note that the correlation function $\overline{c_\ell C^{(n-1)}}$ satisfies the same linear equation as $\overline{C^{(n)}}$ does (cf. Equation 12), which we abbreviate as

$$\overline{\partial_t c_\ell(x_1; t) C^{(n-1)}(x_2, \dots, x_n; t)} = \mathcal{S} \overline{c_\ell C^{(n-1)}} \quad (\text{B9})$$

$$\partial_t \overline{C^{(n)}}(x_1, \dots, x_n; t) = \mathcal{S} \overline{C^{(n)}}, \quad (\text{B10})$$

defining a linear (integro-differential) operator $\mathcal{S}$. Imagine solving for the left eigenvector of $\mathcal{S}$ corresponding to eigenvalue 0,

$$0 = \mathcal{S}^\dagger M^{(n)}, \quad (\text{B11})$$

where $M^{(n)}(x_1, \dots, x_n)$ is a function of n variables just like $C^{(n)}$. Then, contracting Equation B8 with this new function yields

$$\langle M^{(n)} | \overline{c_\ell C^{(n-1)}} \rangle = \text{const.} = p^{\text{fix}} \langle M^{(n)} | \overline{C^{(n)}} \rangle. \quad (\text{B12})$$

The second equality is a key step. It holds because on long times only two outcomes are possible in the stochastic dynamics for $c_\ell C^{(n-1)}$: Either $c_\ell C^{(n-1)} \rightarrow C^{(n)}$ if and only if the subpopulation c_ℓ fixates, or otherwise $c_\ell C^{(n-1)} \rightarrow 0$ when it goes extinct.

Hence, the fixation probability p_ℓ^{fix} of the labeled subpopulation with initial density $c_{\ell,0}(x)$ at time 0 is given by

$$p_\ell^{\text{fix}} = \frac{\langle M^{(n)} | \overline{c_\ell C^{(n-1)}} \rangle}{\langle M^{(n)} | \overline{C^{(n)}} \rangle}. \quad (\text{B13})$$

Fortunately, the left eigenvector of $\mathcal{L}$ is easily constructed: If we fully contract Equation B10 with n factors of the weighting function $u^{(n)}(x_i)$, we have to get 0 on the left-hand side because of the constraint Equation 7. Hence, the sought-after left eigenvector can be written as

$$\langle M^{(n)} | \overline{C^{(n)}} \rangle = \left(\prod_i \int_{x_i} u^{(n)}(x_i) \right) \overline{C^{(n)}}(x_1, \dots, x_n). \quad (\text{B14})$$

Since this eigenvectors contracts to 1 with the total population, Equation B13 becomes

$$p_\ell^{\text{fix}} = \langle u^{(n)} | c_{\ell,0} \rangle = \frac{2}{b(n+1)} \int_x w(x) c_{\ell,0}(x). \quad (\text{B15})$$

Thus, as announced in *Summary of Main Formal Results*, a single labeled mutant at a certain location x has probability $u^{(n)}(x)$ that its descendants will take over the population on long times.

Decay of heterozygosity

The diversity of labels in a population will inevitably decline, because of the rise and ultimate fixation of one of the labels initially present. One can capture the gradual loss of genetic diversity by studying the expectation of the product of density fields that correspond to different types. For instance, if there are just two types, $i \in \{1, 2\}$, the correlation function $\overline{C}_{12}^{(2)}(x, y; t) = \overline{c_1(x, t) c_2(y, t)}$ satisfies

$$\partial_t \overline{C}_{12}^{(2)}(x, y; t) = \sum_{j=1}^{n=2} \left(\mathcal{L} + \gamma(t) - \frac{2n}{n+1} w \right) \Big|_j \overline{C}_{12}^{(2)}, \quad n = 2. \quad (\text{B16})$$

Note that the right-hand side of Equation B16 misses the positive δ -function term of Equation 12, which originated from correlations of subpopulations with themselves. Assuming that $\overline{C}^{(2)}$ as well as $\overline{C}_{1,1}^{(2)}$ and $\overline{C}_{2,2}^{(2)}$ have stable stationary solutions, we can conclude that $\overline{C}_{1,2}^{(2)}$ will decay to 0 at long times because of the lacking source term. This is expected because the gradual fixation of one of the two types implies that $\overline{C}_{1,2}^{(2)}$ has to approach 0 on long times. Discerning relaxation times of the time-evolution (B16) is related to a key question in population genetics: How fast does genetic diversity decay over time?

The function $\overline{C}_{1,2}^{(2)}$ is closely related to the so-called heterozygosity in population genetics. The heterozygosity in the population is the probability that two randomly chosen individuals are of different type,

$$H(t) = \int_x \int_y \overline{C}_{1,2}^{(2)}(x, y; t) / N^2(t). \quad (\text{B17})$$

For our purposes, it is much more convenient and natural to consider a variant of that quantity,

$$\begin{aligned} \hat{H}(t) &= \int_x \int_y u^{(n)}(x, t) u^{(n)}(y, t) C_{1,2}^{(2)}(x, y; t) \\ &= \langle u^{(n)} | c_1 \rangle \langle u^{(n)} | c_2 \rangle \\ &= p(t) q(t) \end{aligned} \quad (\text{B18})$$

with $p(t) \equiv \langle u^{(n)} | c_1 \rangle$ and $q(t) \equiv \langle u^{(n)} | c_2 \rangle$. These quantities have convenient mathematical properties and natural interpretations. At any time, both $p(t)$ and $q(t)$ represent the probability of fixation of the respective subpopulations. Accordingly, we have $p(t) + q(t) = 1$, ensured by the global constraint. Thus, $\hat{H}(t)$ is similar to a heterozygosity in neutral populations and decays at the same rate as the actual heterozygosity (on long times).

Figure 7 Decay of heterozygosity in the staircase mutation kernel. The staircase mutation kernel is defined by the fact that all mutations have the same effect s occurring at a rate μ (see Equation 3). (A) In stochastic simulations, the heterozygosity $\widehat{H}(t)$, cf. Equation B18, clearly decays exponentially on long times. For very small mutation rates ($\mu/s = 10^{-3}$) the slow turnover of new fitness classes imprints its dynamics onto most (time-dependent) observables, as can be observed by the oscillations on top of the exponential decay. (B) The measured (exponential) decay rates are compared with the approximation $\tau_{\widehat{H},\text{approx}} \propto \langle w^2 | \bar{c}_t \rangle^{-1}$; cf. Equation B21. The agreement between measured and predicted timescales shows that the timescale separation underlying our prediction should also hold for different mutation kernels.

An equation of motion for the expectation $\overline{\widehat{H}(t)}$ can be derived by contracting Equation B16 twice with $u^{(n)}$ and using the equation of motion of $u^{(n)}$, Equation 11,

$$\partial_t \overline{\widehat{H}(t)} = -\frac{2}{b} \frac{n-1}{(n+1)^2} \left(\overline{p(t) \langle w^2 | c_2 \rangle} + \overline{q(t) \langle w^2 | c_1 \rangle} \right). \quad (\text{B19})$$

Note that the right-hand side is always negative. Thus, again, we see that the heterozygosity will necessarily decay.

Time-scale separation: As discussed in the main text, noisy traveling waves seem to be characterized generically by a timescale separation: For large $\ln N$, the longest relaxation time of the mean density field is shorter than the typical decay time of the genetic diversity. In the case of adaptation waves and invasion waves with a diffusion kernel, this is evident from the gap in the two lowest relaxation times in our numerical results in Figure 5 and Figure 6B (also see Figure B2 which shows the full spectrum of the operator obtained numerically).

In the presence of such a timescale separation, we can approximate $c_1(x, t) \approx p(t)\bar{c}(x, t)$ and $c_2(x, t) \approx q(t)\bar{c}(x, t)$. In words, we assume that the dynamics on timescales $\tau_n \ll \tau_0$ distribute the subpopulations over the whole extent of the population long before one of them fixates. For the purpose of using these approximations in the terms involving $\langle w^2 | c_i \rangle$ in Equation B19, we need them to be accurate in the high-fitness tail. Using $\overline{\widehat{H}(t)} = \overline{p(t)q(t)}$, we obtain

$$\partial_t \overline{\widehat{H}(t)} \approx -\frac{4}{b} \frac{n-1}{(n+1)^2} \langle w^2 | \bar{c} \rangle \overline{\widehat{H}(t)}, \quad (\text{B20})$$

suggesting again that the leading decay timescale of the heterozygosity can be approximated by

$$\tau_{\widehat{H},\text{approx}} \approx \frac{b(n+1)^2}{4(n-1)} \langle w^2 | \bar{c} \rangle^{-1}. \quad (\text{B21})$$

This approximation indeed seems to approach the correct timescale for large $\ln N$, as can be appreciated from Figure 5 and Figure 6B.

To further test the hypothesis in Equation B21 we conducted simulations as well as numerical estimations for the staircase mutation kernel (see Equation 3, all mutations have the same effect s). Results are summarized in Figure 7, which shows a good agreement of the approximation compared to measured simulation data.

Separation ansatz: As in Equation B16 for $n = 2$, one can easily see that the equation of motion for $\bar{C}_{i_1, \dots, i_n}^{(n)}$ is separable for the n components if $i_j \neq i_l$. It therefore admits a solution of the simple form

$$C_{i_1, \dots, i_n}^{(n)} = \prod_{j=1}^n f_{i_j}(x_j; t), \quad (\text{B22})$$

with $f_{i_j}(x_j; t)$ satisfying a one-dimensional linear equation

$$\partial_t f_{ij}(x; t) = \left(\mathcal{L} + \gamma(t) - \frac{2n}{n+1} w \right) f_{ij}(x; t), \quad (\text{B23})$$

subject to the initial conditions $f_{ij}(x; 0) = c_{ij}(x; 0)$.

Note that contracting with w and using its dynamics, $-\partial_t w = (\mathcal{L}^\dagger + \gamma - w)w$, results in

$$\partial_t \langle w | f_{ij} \rangle = -\frac{n-1}{n+1} \langle w^2 | f_{ij} \rangle. \quad (\text{B24})$$

Thus, $\langle w | f(x; t) \rangle$ must be continuously decaying with time for $n > 1$.

Due to the one-dimensional nature of Equation B22, it is possible to obtain good approximations to the longest relaxation time for various models, by asymptotically solving the associated eigenvalue problem. Below, we provide such a calculation for the adaptation waves with a diffusive kernel in the limit of large populations (or fast waves).

Decay of genetic diversity in models of adaptation: Scaling of the longest relaxation time: Here, we show that the decay of genetic diversity in models of adaptation with a diffusion kernel is related, in a simple way, to the relaxation times of the fixation probability in the limit of large wave speeds. Among other things, this allows us to obtain another simple asymptotic prediction for longest decay rate of the heterozygosity that does not rely on knowing the full profiles in fitness space as in Equation B21. However, this prediction uses the diffusion kernel for mutations. Hence it does not generalize to the other mutation models in Equation 3, in contrast to the approximation derived above.

We proceed as follows. We compare Equation 15 for mixed correlation functions, encoding the decay of genetic diversity, with Equation 11 of the weighting function $w(x; t)$, which is directly proportional to the fixation probability, Equation 10. Since both linear differential equations are very similar, we ask how eigenvalues and eigenfunctions of one equation can be obtained from the ones of the other equation. This is useful because we know that the dynamics of the weighting function have a zero eigenmode, corresponding to the steady-state fixation probability. This eigenmode can then be translated into a slowly decaying mode of the heterozygosity with a characteristic decay rate, which is the main quantity we are looking for.

Before we begin this analysis, we note that the Liouville operator in the comoving frame,

$$(\mathcal{L}_x f)(x) = D \partial_x^2 f + x f + v \partial_x f, \quad (\text{B25})$$

is not self-adjoint, $\mathcal{L}^\dagger \neq \mathcal{L}$. However, the transformation

$$f(x) = \exp\left(-\frac{vx}{2D}\right) \phi(x) \quad (\text{B26})$$

turns Equation B25 into

$$(\mathcal{L}_x f)(x) = \exp\left(-\frac{vx}{2D}\right) \left(\tilde{\mathcal{L}}_x \phi \right)(x) \quad (\text{B27})$$

in terms of a transformed linear operator

$$\tilde{\mathcal{L}} = D \partial_x^2 - \frac{v^2}{4D} + x. \quad (\text{B28})$$

The advantage of carrying out this variable transformation is that $\tilde{\mathcal{L}} = \tilde{\mathcal{L}}^\dagger$ is self-adjoint. Note that this key transformation to a self-adjoint operator relies on the jump kernel being a simple diffusion operator.

The separation ansatz from the previous paragraph showed that the mixed correlation function of n labeled subpopulations requires solving the one-dimensional partial differential Equation B22 for the function $f_{ij}(x_j, t)$, where i_j is the index of the j th labeled subpopulation. To simplify the notation we suppress the index in the following calculation. Then, after substituting $f(x, t)$ in terms of the transformed function $\phi(x, t) \equiv f(x, t) \exp(vx/2D)$, we obtain

$$\partial_t \phi^{(n)}(x, t) = \left[\tilde{\mathcal{L}} - \frac{2n}{n+1} w \right] \phi^{(n)}, \quad (\text{B29})$$

where the superscript $^{(n)}$ is used to distinguish solutions for models that are tuned to be closed at different moments n . Inserting Equation B28, this can also be written as

$$\partial_t \phi^{(n)}(x, t) = \left[D \partial_x^2 - U^{(n)}(x) \right] \phi^{(n)}, \quad (\text{B30})$$

Figure B1 Relaxation rate spectra for models of adaptation and invasion. Shown is quantification of the relaxation rates of correlations among labeled subpopulations in models of adaptation (A–C) and invasion (E–G). The relaxation rates λ_i were found from a spectral analysis of Equation 15. All eigenvalues are real and negative; i.e., all correlations decay with time. Each λ_i corresponds to a relaxation time $\tau_i = -1/\lambda_i$. D and H show the behavior of the two longest timescales τ_0 and τ_1 . The ratio τ_0/τ_1

controls the timescale separation between coalescence (slow) and wave profile relaxation (fast). In addition, the effects of choosing a particular n for the closure of the moment hierarchy vanish for large population sizes (or wave speeds). Relaxation timescales of eigenmodes become more and more similar. Figure 5 and Figure 6 depict only the first two timescales τ_0 and τ_1 (here shown as solid circles in A–C and E–G and emphasized again in D and H) and concentrate on the case $n = 2$.

where we defined the potential function

$$U^{(n)}(x) = \frac{v^2}{4D} - x + \frac{2n}{n+1}w(x). \quad (\text{B31})$$

Importantly, when the wave speed is large, we can relate the potential functions corresponding to different n in a very simple way,

$$U^{(n)}(x) \approx U^{(1)}(x + \Delta x^{(n)}) + \Delta x^{(n)}, \quad (\text{B32})$$

with a shift

$$\Delta x^{(n)} = \frac{2D}{v} \ln \left(\frac{2n}{n+1} \right). \quad (\text{B33})$$

This reformulation is possible when the wave speed is high, because $w(x) = \psi(x)\exp(vx/2D)$ is the product of a rapidly varying exponential and, at least close to its peak, a slowly varying function $\psi(x)$. The shape of $\psi(x)$ is shown in Figure B2, A and B. Thus,

$$w(x + \Delta x^{(n)}) \approx w(x) \exp \left[\frac{v\Delta x^{(n)}}{2D} \right]. \quad (\text{B34})$$

Choosing a different n results in a shift by $\Delta x^{(n)}$ of the potential in both directions, vertically and horizontally.

An equation of the type of Equation B30 is formally equivalent to a Schrödinger equation, which describes the dynamics (in imaginary time) of a quantum particle in a “potential” $U^{(n)}(x)$. This analogy is mathematically useful because one can benefit from the rich body of knowledge about Schrödinger equations.

Using our numerical solutions for $w(x)$, we can plot in Figure B1A the actual form of the potential for adaption waves and test the scaling property hypothesized in Equation B32. Note that the potential has the shape of a well, which confines the quantum mechanical particle as long as its “energy” is small enough.

A confined quantum mechanical particle is characterized by a *discrete* set of eigenstates. This means that there is a discrete set $\{\phi_k^{(n)}(x)\}$ of eigenfunctions and $\{\lambda_k^{(n)}\}$ of eigenvalues, labeled by a discrete index k , which corresponds to the localized states of the quantum particle. The lowest eigenfunction is indicated in Figure B1.

The remaining eigenvalue problem thus reads

$$\lambda_k^{(n)} \phi_k^{(n)}(x) = [D\partial_x^2 - U^{(n)}(x)] \phi_k^{(n)}. \quad (\text{B35})$$

Inserting the scaling property Equation B32 for the potential function, and using shifted variables $y \equiv x + \Delta x^{(n)}$, yields

$$(\lambda_k^{(n)} + \Delta x^{(n)}) \phi_k^{(n)}(y - \Delta x^{(n)}) = [D\partial_y^2 - U^{(1)}(y)] \phi_k^{(n)}(y - \Delta x^{(n)}). \quad (\text{B36})$$

By comparing Equations B35 and B36, we arrive at a key result:

$$\lambda_k^{(n)} \approx \lambda_k^{(1)} - \Delta x^{(n)}, \quad (\text{B37})$$

$$\phi_k^{(n)}(x, t) \approx \phi_k^{(1)}(x + \Delta x^{(n)}, t). \quad (\text{B38})$$

The k th eigenfunction of the operator $\tilde{\mathcal{L}}^{(n)} \equiv D\partial_x^2 + U^{(n)}(x)$ is given by the k th eigenfunction of $\tilde{\mathcal{L}}^{(1)}$ shifted by $\Delta x^{(n)}$ toward lower fitness and with eigenvalues shifted by the same amount $\Delta x^{(n)}$.

Moreover, the operator on the right-hand side of Equation B36 also appears in Equation 11 for the weighting function $w(x; t)$. Therefore, we know that this operator admits a steady-state

$$0 = \left[D\partial_y^2 - U^{(1)}(y) \right] w(x). \quad (\text{B39})$$

Thus, we can conclude, in particular, for the lowest eigenmode (the “ground state”),

$$\lambda_0^{(n)} \approx \Delta x^{(n)}, \quad (\text{B40})$$

$$\phi_0^{(n)}(x, t) \approx w\left(x + \Delta x^{(n)}\right) \exp\left(-\frac{\nu}{2D}\left(x + \Delta x^{(n)}\right)\right). \quad (\text{B41})$$

This analysis suggests that the heterozygosity in the n -tuned model of adaptation with a diffusion kernel decays as $\exp(-2\Delta x^{(n)}t)$ on long times. A test of this assertion is shown in Figure B2, B and C. In general, m -point correlation functions decay like $\exp(-m\Delta x^{(n)}t)$, asymptotically.

For intermediate times, the higher eigenvalues matter as well. These are spaced quite evenly, as seen in our numerical determined relaxation rate spectra of Figure B2. Qualitatively, this feature can be understood again through the quantum particle analogy. The quantum particle is confined to a potential of width of order $L_U = O(D^{1/3})$. Approximating the potential well by a square potential well, we indeed expect evenly spaced energy levels of order $\lambda_k^{(n)} \sim kd/L_U^2 \sim O(kD^{2/3})$. However, because the potential well is instead wedge shaped (see Figure B1), the spacing decreases with increasing index k , as can be appreciated in the numerically measured spectrum of Figure B2.

Figure B2 Potential energy function and decay of genetic diversity in models of adaptation. A and B show the potential energy functions $U^{(1)}(x)$ and $U^{(2)}(x)$ for the two wave speeds of $vD^{-2/3} = 5, 10$ in dark and light blue, respectively. The approximation by the shifted function $U^{(2)}(x) \approx U^{(1)}(x + \Delta x^{(2)}) + \Delta x^{(2)}$ is also shown in gray and agrees well with $U^{(2)}(x)$ in the region where the lowest eigenfunction $\phi_0^{(1)}(x)$ (green, in arbitrary units) has most of its support. Note that the lowest eigenfunction $\phi_0^{(1)}(x)$ is identical to the transformed weighting function $\psi(x) \equiv w(x)\exp(-vx/2D)$. (C) Decay of heterozygosity as a function of time for a set of velocities $vD^{-2/3} = \{1, \dots, 10\}$ is exponential on long times. The extracted decay rate is plotted along with other predictions in Figure 5. Reassuringly, the measured decay rates follow the predicted $v^{-1} \sim (\ln N)^{-1/3}$ scaling. Measurements have been conducted in simulations set up similar to those used in Figures 3 and 10. The shown heterozygosity $\hat{H}$ (cf. Equation B18) is tracked and averaged over 2000 simulation runs.

Analogous considerations can be done for the decay of genetic diversity in other traveling-wave models, for instance for invasion waves, as long as the jump kernel is diffusive.

Appendix C: Explicit Equations of Motion in Tuned Models

In the main text, we presented general equations for the correlation functions $\bar{C}^{(n)}$ and the weighting function w . Numerical results have been computed for the two cases $n = 1$ and $n = 2$. For reference, here we state the explicit equations of motion for both, adaptive and invasive, waves.

Adaptation waves

The first step is always to calculate the weighting function w in a comoving frame with speed v :

$$0 = -v\partial_x w(x) + D\partial_x^2 w(x) + xw(x) - w(x)^2. \quad (C1)$$

Here, the term with multiplication of fitness, $xw(x)$, indicates the selection term in adaptive waves, which has to be replaced with $sH(x)w(x)$ for invasive waves (see below).

Model tuned to the first moment: After obtaining the general form of the weighting function w , the mean stationary population density follows as

$$0 = v\partial_x \bar{C}^{(1)}(x) + D\partial_x^2 \bar{C}^{(1)}(x) + x\bar{C}^{(1)}(x) - w(x)\bar{C}^{(1)}(x), \quad (C2)$$

a result that has already been described in Hallatschek (2011).

Model tuned to the second moment: If we choose $n = 2$, the model is tuned to have a closed two-point correlation function which is governed by

$$\begin{aligned} 0 = & v\partial_x \bar{C}^{(2)}(x, y) + v\partial_y \bar{C}^{(2)}(x, y) + D\partial_x^2 \bar{C}^{(2)}(x, y) + D\partial_y^2 \bar{C}^{(2)}(x, y) \\ & + (x + y - 4w(x)/3 - 4w(y)/3)\bar{C}^{(2)}(x, y) + 2/3 \delta(x - y) \int dz w(z)\bar{C}^{(2)}(x, z). \end{aligned} \quad (C3)$$

In this case, the mean stationary population density is obtained by contraction, $\bar{C}^{(1)} = \langle u^{(2)} | \bar{C}^{(2)} \rangle$. Note, however, that this expression is not identical to the density $\bar{C}^{(1)}$ in the model tuned for the first moment, $n = 1$. This can be appreciated by applying $\langle w | \cdot \rangle$ to Equation C3, which yields

$$0 = v\partial_x \langle w | \bar{C}^{(2)} \rangle + (\mathcal{L} - 2/3w) \langle w | \bar{C}^{(2)} \rangle - 1/3 \langle w^2 | \bar{C}^{(2)} \rangle, \quad (C4)$$

as dynamics of the mean density in the model with $n = 2$. Here, the problem of stochastic nonlinear dynamics resurfaces again in the last term: The hierarchy of moment equations is not closed in general. Even in our tuned approach *only* the equation with the chosen n decouples from higher-order correlation functions. Still, as we have shown in Figure 2 the contraction of the two-point correlation function to obtain the mean density leads to very similar profiles for the two models tuned for $n = 1$ and $n = 2$, respectively.

Invasion waves

For invasion waves the weighting function w is the solution to

$$0 = -v\partial_x w(x) + D\partial_x^2 w(x) + sH(x)w(x) - w(x)^2, \quad (C5)$$

which again sets the speed v of the comoving frame as a parameter. In addition, tuning the model for the first moment, $n = 1$, yields the equation of motion for the stationary mean population density $\bar{C}^{(1)}(x)$,

$$0 = v\partial_x \bar{C}^{(1)}(x) + D\partial_x^2 \bar{C}^{(1)}(x) + sH(x)\bar{C}^{(1)}(x) - w(x)\bar{C}^{(1)}(x). \quad (C6)$$

Choosing $n = 2$ for invasion waves leads to

$$\begin{aligned} 0 = & v\partial_x \bar{C}^{(2)}(x, y) + v\partial_y \bar{C}^{(2)}(x, y) + D\partial_x^2 \bar{C}^{(2)}(x, y) + D\partial_y^2 \bar{C}^{(2)}(x, y) \\ & + (sH(x) + sH(y) - 4w(x)/3 - 4w(y)/3)\bar{C}^{(2)}(x, y) + 2/3 \delta(x - y) \int dz w(z)\bar{C}^{(2)}(x, z). \end{aligned} \quad (C7)$$

Appendix D: Numerical Methods

Although asymptotic analyses of noisy traveling-wave models are often possible, closed-form solutions of either the fixation probability w or correlation function $\bar{C}^{(n)}$ are usually out of reach. Numerical methods can help alleviate this problem, as with tuned models we are able to state at least *exact* moment equations, which do not need any further approximations or assumptions.

To solve the governing equations of our tuned model numerically, we implemented an algorithm based on a multidimensional Newton–Raphson (NR) iteration scheme (Press *et al.* 2007; Geyrhofer 2014). Here, we recount the basic steps of this scheme.

For the strictly one-dimensional case, $n = 1$, the numerical solution involves the roots of a set of M equations in M variables,

$$f_i(y_1, \dots, y_M) = 0, \quad 1 \leq i \leq M, \quad (D1)$$

which represent the steady-state equations of motion, e.g., Equations C1 and C2 for the case of adaptation and $n = 1$. The variables y_i denote the value of the desired weighting function or correlation function at lattice point i , respectively.

For a small deviation δy_i in one of the variables, one can expand f_i into a series,

$$f_i(y_1, \dots, y_i + \delta y_i, \dots, y_M) = f_i(y) + \partial_{y_i} f_i(y) \delta y_i + O(\delta y_i^2). \quad (D2)$$

In the NR scheme, one iterates the current guess for the solution y^{old} to obtain $y^{\text{new}} = y^{\text{old}} + \delta y$. The small difference δy is extrapolated by assuming the new solution y^{new} fulfills the equation of motion, $f_i(y^{\text{new}}) = 0$, while truncating (D2) after the linear term. This leads to the equation $0 = f_i(y^{\text{old}}) + \partial_{y_i} f_i(y^{\text{old}}) \delta y_i$. Thus, a single step composes evaluating the expression

$$y_i^{\text{new}} = y_i^{\text{old}} - \frac{f_i(y_{i-1}^{\text{old}}, y_i^{\text{old}}, y_{i+1}^{\text{old}})}{\partial_{y_i} f_i(y_{i-1}^{\text{old}}, y_i^{\text{old}}, y_{i+1}^{\text{old}})} \quad (D3)$$

for each lattice point. To ensure a better convergence, the solutions y_i^{new} for even and odd indexes are computed consecutively. In the notation of (D3) we also made use of a simplifying fact: The diffusion approximation (for the mutation process in the adaptive wave and movement in the invasive waves) leads only to a “local” coupling, such that the equations of motion depend only on the values of y at the focal lattice site i and the two neighboring ones, $f_i(y_{i-1}, y_i, y_{i+1})$. For the case of adaptation waves, we discretize Equation C1 and obtain the required expressions for the weighting function w ,

$$f_i(w_{i-1}, w_i, w_{i+1}) = (D/dx^2 + v/2dx)w_{i-1} + (x_i - 2D/dx^2)w_i + (D/dx^2 - v/2dx)w_{i+1} - w_i^2, \quad (D4)$$

$$\partial_{w_i} f_i(w_{i-1}, w_i, w_{i+1}) = (x_i - 2D/dx^2) - 2w_i, \quad (D5)$$

which have to be inserted into Equation D3. For obtaining solutions, the lattice spacing dx has to be chosen small enough that the right-hand side of (D5) is negative on the whole lattice and does not change sign for any $x_i = i dx$. The range of i has to be adjusted to fit all characteristic features of the profiles onto the M lattice points. Similar expressions to (D4) and (D5) hold for the mean stationary population density $\bar{C}^{(1)}$ after discretizing (C2).

Boundary conditions for the numerical algorithm are usually chosen to represent an exponential decay. Thus, the value at a point $i - 1$ preceding the lattice is set to y_i^2/y_{i+1} , while the next point following the lattice $i + 1$ is set to y_i^2/y_{i-1} . In contrast, the upper boundary $i + 1$ for the weighting function is assumed to be a linear extrapolation, $2y_i - y_{i-1}$. Initial conditions for the weighting function are usually $u_i \approx x_i/2$ for $x_i \geq 0$ and setting it to a small value, e.g., $u_i \approx dx$ for $x_i < 0$. Using more elaborate semianalytical expressions speeds up the time to convergence. For the population density, a Gaussian approximation with variance v , $c_i \sim \exp(-x_i^2/2v)$, suffices as an initial condition. The correct coefficient is ensured by repeated rescaling to comply with the constraint, $\sum_i u_i c_i = 1/dx$.

For the higher-dimensional correlation functions $\bar{C}^{(n)}$, $n > 1$, the method can be extended in a straightforward fashion. For instance, for $n = 2$ one has $M \times M$ variables y_{ij} and $M \times M$ functions f_{ij} . Each dimension adds only two additional variables in the equation of motion, $f_{ij}(y_{i-1,j}, y_{i,j-1}, y_{i,j}, y_{i+1,j}, y_{i,j+1})$. Note that the discrete limit of the Dirac delta for correlations in the last term of Equation C3 is given by $\delta(x_i - x_j) = 1/dx \delta_{ij}$.

An improvement of this algorithm that utilizes not only the local derivative $\partial_{y_i} f_i$, but also the whole Jacobian with entries $\mathcal{J}_{ij} = \partial_{y_j} f_i$, is often needed for extended mutation kernels $\mu(y)$ in Equation 3 (Geyrhofer 2014; Geyrhofer and Hallatschek, unpublished results). In these cases (and for reasonable parameter values), Equation D5 changes its sign twice on any lattice,

regardless of the choice of lattice spacing dx , which renders this local approximation (D3) unusable. However, (D3) suffices for all present purposes.

In Figure B1 we display the spectral decomposition of the linear operator governing the equation of motion for the stationary mean population density. In this case, the governing (discretized) equations (D1) are given by the linear equation $0 = f_i(y_1, \dots, y_M) = \sum_j F_{ij} y_j$, with coefficients F_{ij} obtained from discretizing Equation 15. The eigenvalues λ_i are obtained by a Schur decomposition of the matrix F_{ij} , which leads to an upper or quasi-upper triangular matrix: Along its diagonal it has 1×1 blocks with real eigenvalues and 2×2 blocks with its complex conjugate eigenvalues. The existence of complex eigenvalues depends on the mutation (or migration) scheme. For a diffusion scheme (*i.e.*, a second derivative) in Equation 3 we obtain real eigenvalues only numerically. The code itself is based on the already implemented routines in the GNU Scientific Library (GSL), in particular centered around the function `gsl_eigen_nonsymm` to provide input and parse output.

The numerical code is available at <https://github.com/lukasgeyrhofer/adaptivewaves>.

Appendix E: Stochastic Simulations

For the stochastic simulations of adaptation waves, the continuous density $c_t(x)$ in fitness space is discretized into bins on a regular one-dimensional lattice. All individuals in the interval $[x_i, x_{i+1}]$ with $x_i = i dx$ are counted in the *occupancy vector* n_i ,

$$n_i = c_t(x_i) dx. \quad (E1)$$

The occupancies are updated in discrete time steps of length ϵ , which encompass the dynamics in the stochastic Equation 1 and the subsequent step to limit population sizes, Equation 5 or 7. The action of these equations consists of three substeps in the algorithm, indicated by superscripts in subsequent equations. In each of those steps the occupancies (can) change.

First, the mean (deterministic) change due to a comoving frame, mutations, and selection is applied,

$$\Delta^{(1)} n_i / \epsilon = \frac{v}{2dx} (n_{i+1} - n_{i-1}) + \frac{D}{dx^2} (n_{i-1} - 2n_i + n_{i+1}) + (x_i - x_0(t)) n_i. \quad (E2)$$

The first term for the comoving frame is used only in simulations with a tuned constraint. The next term represents modifications in fitness due to mutations, while the last term represents growth due to selection. For the latter, we have to distinguish again a fixed population size constraint and our tuned constraint. The offset $x_0(t)$ in the selection term is either set to the mean fitness $x_0(t) = \sum_i x_i n_i / N$ in the fixed population size constraint or set to $x_0(t) = 0$ as we incorporated the change in mean fitness already with the comoving frame. For the case of invasion waves, the selection term is replaced by $sH(x_i - x_0(t)) n_i$. There the offset $x_0(t)$ is the position of the front, which we define as $x_0(t) = \sum_i x_i n_i n_{-i} / \sum_i n_i n_{-i}$ for a fixed population size constraint. Again, we have $x_0(t) = 0$ for the tuned constraint in its comoving frame.

In the next substep, the randomness due to birth and death events (*i.e.*, genetic drift) further modifies all occupancies n_i ,

$$\Delta^{(2)} n_i / \sqrt{\epsilon} = \sqrt{2} (\text{Poisson}(n_i) - n_i), \quad (E3)$$

where $\text{Poisson}(n_i)$ is a Poisson-distributed random number with parameter n_i . This particular form of the noise (with n_i already updated from mutations and selection) ensures that (i) occupancies n_i do not drop below zero, (ii) the mean value of the noise is zero, and (iii) the variance at a lattice site i amounts to $2\epsilon n_i$ per time step ϵ . While the introduction of occupancies n_i instead of a density c_i is irrelevant for most of the algorithm, it is convenient for this last feature (iii). The correlations of the noise (Equation 4) are given by the expression $\overline{\eta_t(i dx) \eta_{t'}(j dx)} = \delta_{tt'} (1/dx \delta_{ij})$ in the discretized simulations, where the factor $1/dx$ is then scaled already into the n_i .

In the last substep, the population is scaled uniformly to comply with its constraint,

$$\Delta^{(3)} n_i = \left(\frac{1}{\sum_j u_j n_j} - 1 \right) n_i. \quad (E4)$$

For the fixed population size constraint in Appendix B, we simply set $u_j = N^{-1}$ for all j [or, for invasion waves $u_j = N^{-1}$ for $j dx > x_0(t)$ and $u_j = 0$ otherwise]. In the case of tuned models, we first solve the equation of motion for the fixation probability in a comoving frame. To obtain this numerical solution for w , we utilize the code presented in Appendix D.

Simulation code is available at <https://github.com/lukasgeyrhofer/adaptivewaves>.

Figure F1 (A) Fixation and extinction measurements with the trivial labeling of $c_t(x, 0) = c(x, 0)/2$. With such a labeling we obviously expect half of all simulations to end in fixation and half to end in extinction of the subpopulation c_t . The obtained confidence intervals serve as an indicator for the noisiness of such measurements, when comparing results to the position-dependent labeling (cf. Equation F2) presented in Figure 3. In general, however, the initial hypothesis of a fixation probability of $1/2$ is corroborated by simulation results (B and C).

Appendix F: Measuring Fixation Probabilities

In Figure 3, B and C, we presented confidence intervals for the fixation probability, obtained via measurements of fixation and extinction events. *A priori*, counting events leads to an average value for the fixation probability, which might or might not be close to the expected (theoretical) value. To compute confidence intervals, additional assumptions have to be made, explained below.

After generating different starting conditions c_t (snapshots from stochastic simulations), we label subpopulations c_t in the nose of the wave, such that

$$\langle u^{(n)} | c_t \rangle = \frac{1}{2}. \quad (\text{F1})$$

There are, of course, many ways to label a subpopulation such that the expected fixation probability is 50%. The simplest way is to label 50% of the population in each bin, which unsurprisingly yields fixation probabilities very close to 50% (cf. Figure F1). However, this naive labeling protocol does not test our predictions for the spatial dependence of the fixation probability. To test the accuracy of our predictions in the spatially varying region of the fixation probability, we label the population in the tip of the wave, using the following form:

$$c_t(x; t) = \frac{c(x; t)}{1 + \exp(-(x - x_t)/\delta)}. \quad (\text{F2})$$

Here, x_t determines the position of the labeling and δ its steepness. To avoid artifacts associated with the discreteness, we choose the length scale δ such that on the order of 10 bins typically contain both labeled and unlabeled subpopulations. The crossover x_t is iteratively adjusted until the condition in Equation F1 is met with sufficient accuracy (usually, we demand a value close to machine precision, 10^{-10}).

From this configuration, we run the stochastic time evolution of the population until the labeled subpopulation either reaches fixation or goes extinct. In accordance with our interpretation of $u^{(n)}$ as fixation probability, we expect the labeled population to fix in half of the simulation runs and to go extinct otherwise. For definiteness, we abort simulations when one of the thresholds $\langle u^{(n)} | c_t \rangle = 1 - 10^{-4}$ or $\langle u^{(n)} | c_t \rangle = 10^{-4}$ is exceeded. We count such events as fixation and extinction, respectively.

After having amassed such simulation evidence, we can use Bayesian inference to check whether our assumption of the interpretation of $u^{(n)}$ as fixation probability is consistent (O'Hagan and Forster 1994). The distribution of the fixation probability p_t^{fix} of the subpopulation, given the (simulation) data is computed as

$$\mathbb{P}[p_t^{\text{fix}} | \text{data}] \sim \mathbb{P}[\text{data} | p_t^{\text{fix}}] \mathbb{P}[p_t^{\text{fix}}], \quad (\text{F3})$$

using Bayes' theorem. Here, the likelihood $\mathbb{P}[\text{data} | p_t^{\text{fix}}]$ of observing either an extinction or a fixation event is a simple binomial distribution: When there are N trials with X fixation events, the likelihood is

$\mathbb{P}[X \text{ fixation events} | p_t^{\text{fix}}] = \binom{N}{X} (p_t^{\text{fix}})^X (1 - p_t^{\text{fix}})^{N-X}$. Furthermore, we assume a flat prior $\mathbb{P}[p_t^{\text{fix}}]$ for the fixation probability

p_ℓ^{fix} , ignoring any knowledge about its value at the beginning. Such a flat prior can also be cast as a Beta distribution [incidentally a conjugate prior (O'Hagan and Forster 1994)], which in its general form is given by

$$\text{Beta}(Y; \alpha, \beta) \propto \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} Y^{\alpha-1} (1-Y)^{\beta-1}. \quad (\text{F4})$$

Using the two hyperparameters $\alpha = 1$ and $\beta = 1$ we arrive at the uniform (flat) distribution on $[0; 1]$. Thus, the posterior distribution for the fixation probability p_ℓ^{fix} of the subpopulation can also be written as Beta distribution,

$$\mathbb{P}\left[p_\ell^{\text{fix}} \mid X \text{ fixation events}\right] \sim \left(p_\ell^{\text{fix}}\right)^{X+\alpha-1} \left(1-p_\ell^{\text{fix}}\right)^{N-X+\beta-1} \quad (\text{F5})$$

(with $\alpha = \beta = 1$), up to a normalization factor. From this posterior (Equation F5) we can evaluate the 95% confidence interval and check whether our assumption $p_\ell^{\text{fix}} = 1/2$ is within its range. Increasing the value of $\alpha = \beta > 1$ would lead to a peak in the prior of our hypothesis, $p_\ell^{\text{fix}} = 1/2$, thus narrowing the posterior distribution (Equation F5).