PNAS
www.pnas.org

Supplementary Information for

Genealogical structure changes as range expansions transition from pushed to pulled

Gabriel Birzu, Oskar Hallatschek, Kirill S. Korolev

Kirill S. Korolev.
E-mail: korolev@bu.edu

This PDF file includes:

Supporting Information Text

1. Forward in time dynamics

In this section, we show how the genealogy of an expanding population can be mapped to an effective well-mixed population with a broad distribution of reproductive values. We only consider the case of density-independent migration here, but the argument is analogous when D depends on n. As discussed in Ref. (1), we approximate the population front using the steady state solution of the corresponding deterministic equation and incorporate the effects of front fluctuations by using an appropriate cutoff in the low density region at the edge of the front. Using this approximation, the equation describing the front has the form

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} + r(n)n. \quad [1]$$

We assume the population is comprised of m neutral subtypes with relative fractions $f_i(t,x) \equiv n_i(t,x)/n(t,x)$ and $\sum_{i=1}^{m} f_i(t,x) = 1$. In the deterministic limit, it is then easy to show that $f_i(t,x)$ obey the following equation (2, 3):

$$\frac{\partial f_i}{\partial t} = D\frac{\partial^2 f_i}{\partial x^2} + 2D\frac{\partial \ln n}{\partial x}\frac{\partial f_i}{\partial x}. \quad [2]$$

It will be useful in the following to consider the above equation in the comoving frame of reference given by the change of variables $\zeta \equiv x - vt$. Introducing this expression in Eq. (2) we get

$$\frac{\partial f_i}{\partial t} = D\frac{\partial^2 f_i}{\partial \zeta^2} + \left(v + 2D\frac{\partial \ln n}{\partial \zeta}\right)\frac{\partial f_i}{\partial \zeta}, \quad [3]$$

where the time variable t captures the transient time dependence of $f_i(t,\zeta)$ after eliminating translational motion with constant velocity along the x axis. It can be shown that Eq. (3) has the property that it admits a time-invariant quantity of the form

$$\pi = \frac{\int_{-\infty}^{\infty} d\zeta f_i(t,\zeta) n^2(\zeta) e^{v\zeta/D}}{\int_{-\infty}^{\infty} d\zeta n^2(\zeta) e^{v\zeta/D}}. \quad [4]$$

The proof that π is time-invariant can be found in Refs. (2, 3).

In order to understand the meaning of Eq. (4) it is useful to consider the case of a single mutant arising at time $t = 0$ at ζ_i. At this initial time, we can approximate the density of the mutant by a delta function $n_i(0,\zeta) = \delta(\zeta - \zeta_i)$. Denoting the value of π for this specific initial condition by $u(\zeta_i)$, introducing the expression for $n_i(0,\zeta)$ into Eq. (4), and using $n_i(t,x) = f_i(t,\zeta)n(\zeta)$ we find the following expression:

$$\pi|_{n_i(0,\zeta)=\delta(\zeta-\zeta_i)} \equiv u(\zeta_i) = \frac{n(\zeta_i)e^{v\zeta_i/D}}{\int_{-\infty}^{\infty} d\zeta n^2(\zeta) e^{v\zeta/D}}. \quad [5]$$

From the time-invariance of π, it follows that $\lim_{t\to\infty} f_i(t,\zeta) = u(\zeta_i)$. Thus, $u(\zeta_i)$ represents the fraction of the population descended from the original mutant as $t \to \infty$. Note, the above result is strictly valid only for deterministic fronts.

How can we extend the above analysis to stochastic fronts? Here, it is helpful to make an anology with well-mixed models such as the one analyzed in Ref. (4). Consider a population of size N, with discrete generations. Each generation, individuals produce offspring according to a distribution $P(W)$, N of which are chosen to form the next generation. As discussed in Ref. (4), it is necessary to choose $P(W)$ such that $\langle W \rangle > 1$ to ensure that at least N offspring are produced with sufficiently high probability. However, sampling N individuals from the offspring pool each generation ensures that the population size stays constant and that no lineage has an advantage over the rest.

In this simplified model, consider what happens if we sample a subpopulation of size $n_i(0)$ and track the clone size $n_i(\tau)$ over time. At $\tau \to \infty$, the whole population is descended from a single individual and only one clone survives. However, when viewed across an ensemble of populations, we can define the fixation probability of a particular clone u_i as the fraction of ensembles in which clone i is the only surviving clone. Since all individuals are equivalent, u_i is simply proportional to the initial sample size:

$$u_i = \frac{n_i(0)}{N}. \quad [6]$$

Note that in the toy example above, every individual is identical, which is clearly not the case for an expanding population, where individuals closer to the front typically have higher fixation probabilities. However, as we show below, we can reduce the problem of fixation in an expanding population to that of a well-mixed population when the relaxation time τ_m at which $f_i(\tau_m,\zeta) \approx u(\zeta_i)$ is much shorter than the coalescence of the front T_c. Mathematically, τ_m is simply the largest eigenvalue of the right hand side of Eq. (3)*.

*Note that for pulled fronts the two terms in front of $\partial f_i/\partial\zeta$ in Eq. (3) approximately cancel close to the leading edge and the motion of lineages can be approximated by a diffusion over the length of the front, which scales as $\sqrt{D/r_0}\ln N$. This leads to a mixing time $\tau_m \sim r_0^{-1}\ln^2 N$ (5). For pushed fronts there is no known general expression for τ_m to our knowledge.

Thus, after an effective generation time τ_m, a single individual sampled from the front at ζ_i expands into a clone that comprises a fraction $u(\zeta_i)$ of the population of the front. The quantity $u(\zeta_i)$ then defines the long-term reproductive success of the original individual so we will refer to it as the *reproductive value* of the individual. Since spatial information is lost past this point, front dynamics are effectively well-mixed when coarse-grained over timescales $O(\tau_m)$[†].

In analogy with the well-mixed case, we can interpret the reproductive value $u(\zeta_i)$ as the fraction of offspring in the population descended from the ancestor after one generation. The relation between the number of offspring W and u can then be computed using

$$W(\zeta) = N_f u(\zeta), \tag{7}$$

where N_f the population size in the well-mixed region at front. Since N_f is a constant, the distributions of u and W are the same. To simplify the notation, we use the reproductive value as the measure of fecundity throughout.

We can compute the distribution of u by using

$$p(u)du \propto n(\zeta)d\zeta, \tag{8}$$

to eliminate the position ζ. The distribution $p(u)$ will in general depend on the whole shape of the profile, but since we are only interested in the tails of the distribution we only need to consider the region far ahead of the front where $n(\zeta) \sim e^{-k\zeta}$, where $k = \frac{v}{2D}(1 - \sqrt{1 - v_F^2/v^2})$ is a constant (see next subsection). Using this expression we find that $p(u)$ takes the form of a power law

$$p(u) \propto u^{-\frac{2}{1-\sqrt{1-v_F^2/v^2}}} = u^{-2-\alpha}, \tag{9}$$

where

$$\alpha = \frac{2\sqrt{1 - v_F^2/v^2}}{1 - \sqrt{1 - v_F^2/v^2}}. \tag{10}$$

For pulled waves, $v/v_F = 1$ and we have $p(u) \propto u^{-2}$. This distribution has a divergent mean and leads to Bolthausen-Sznitman coalescent (5). In the semi-pushed region, $1 < \frac{v}{v_F} < \frac{3}{2\sqrt{2}}$, and the descendant distribution changes continuously from u^{-2} to u^{-3}. Finally, in fully-pushed waves, it decreases at least as fast as u^{-3}. In this case, the population is described by a Kingman coalescent (8).

A. Deterministic fronts with finite population sizes. Discreteness in real populations requires a modification in the continuum model from Eq. (1). In the case of deterministic fronts one can simply set the density below one individual to zero. This analysis is fairly straightforward in the case of fully pushed and semi-pushed fronts, but some care is required for pulled fronts. For completeness, we briefly outline the argument for deriving these cutoffs below, but interested readers may wish to consult Ref. (9) and Sec. IX of the SI from Ref. (1) for a more extensive discussion.

To simplify the calculations later on we introduce a normalized population density $\rho(\zeta)$, defined as

$$\rho(\zeta) = \frac{n(\zeta)}{N}, \tag{11}$$

where $n(\zeta)$ is the travelling wave solution to Eq. (1) and ζ is the position in the comoving reference frame as defined in the previous subsection. It is easy to show that close to the edge of the expansion, where $\rho \ll 1$, there are two independent solutions for the shape of the front, of the form $e^{-k\zeta}$ and $e^{-q\zeta}$, with

$$\begin{cases} k = & \frac{v}{2D}\left(1 + \sqrt{1 - v_F^2/v^2}\right), \\ q = & \frac{v}{2D}\left(1 - \sqrt{1 - v_F^2/v^2}\right), \end{cases} \tag{12}$$

where the reader may recall the definition of the Fisher velocity given in the main text $v_F = 2\sqrt{r_0 D}$. It can further be shown that in the pushed regime[‡] ($v > v_F$), the front converges to the more sharply decaying solution (see SI, Sec. II of Ref. (1) and (10) for details) given by

$$\rho(\zeta) \sim e^{-k\zeta}. \tag{13}$$

Thus, to find the cutoff distance ζ_d^{pushed} for deterministic fronts we simply invert the equation $\rho(\zeta_d^{\text{pushed}}) = 1/N$, which gives

$$\zeta_d^{\text{pushed}} \approx \frac{1}{k} \ln N. \tag{14}$$

[†] The argument presented above can be made more rigorous, but relies on the equation for the front to obey a deterministic PDE (2). A rigorous extension of these results to stochastic fronts remains an important open question. In order to make progress we use a "tuned model" approach as in Refs. (1, 6, 7), which is elaborated in the next subsection.

[‡] Note that here and in the next subsection we use the term "pushed" to refer to *both* fully pushed and semi-pushed expansions. We do this as a matter of convenience when discussing properties which are same in both fully pushed and semi-pushed regimes.

Note that in general there is an additive correction to Eq. (14), which depends on the details of the model, but not on N.

For pulled fronts the situation is slightly more complicated because the two independent eigenvalues k and q become equal to $k_F = \sqrt{\frac{r_0}{D}}$ when $v = v_F$. A similar problem was addressed in Ref. (9) which analyzed the effect of setting the growth rate $r(\rho)$ to zero below a fixed density $\rho_c = 1/N$ (using our notation). Note this is slightly different from our simulations where the front is discrete and the population density, not the growth rate, is zero below $1/N$. Nevertheless, in the limit of large N we conjecture the two models will give similar results. Close to the leading edge, the solution then should have the form

$$\rho(\zeta) \approx \frac{A\zeta_c}{\pi} \sin\left(\frac{\pi\zeta}{\zeta_c}\right) e^{-k_F\zeta}, \quad [15]$$

where A is an undetermined constant with units of inverse length. We can rewrite this equation in terms of the adimensional lengths

$$L = k_F\zeta, \quad [16]$$

as

$$\rho(L) \approx \frac{A'L_c}{\pi} \sin\left(\frac{\pi L}{L_c}\right) e^{-L}, \quad [17]$$

where all of the variables are now adimensional.

To find the location of the cutoff we look for the solution of the equation $\rho(L) = 1/N$, near L_c. As we will see, the prefactor together with the sine factor in Eq. (17) only contribute a constant shift to the value of L, which can be ignored in the limit of $N \to \infty$ we are interested in. Guided by this insight, we can solve the equation by looking for a solution of the form $L = L_c - \epsilon$. Taking the logarithm on both sides and using the approximation $\sin[\pi(1-x)] \approx \pi x$ for $x \ll 1$ gives

$$L_c - \epsilon - \ln(A'\epsilon) = \ln N. \quad [18]$$

Assuming the dominant balance is between L_c and $\ln N$, we find a self-consistent solution $L_c = \ln N$. The next order correction is then given by the solution to the equation $\epsilon = -\ln(A'\epsilon)$, which will be some constant independent of N. As a check, we indeed have $L_c \gg \epsilon$ for sufficiently large N, as required for our approximation to be valid. Taken together these calculations give the following simple expression for the cutoff in deterministic fronts:

$$\zeta_d^{\text{pulled}} \approx \frac{1}{k_F} \ln N. \quad [19]$$

B. Stochastic fronts with finite population size. For stochastic fronts, the analysis is much more complicated. Nevertheless, theoretical studies starting from Refs. (11, 12) and continuing through Refs. (1, 6) have established that the dynamics of all three classes of expansions can also be described using a similar, but distinct, cutoff in the population density as in the deterministic case. We briefly recapitulate the argument for deriving the cutoff in stochastic fronts below. The presentation closely follows the one from Ref. (1) so readers familiar with those results may wish to skip to the next subsection.

The correct procedure for determining the cutoff for a general growth function $r(n)$ in the case of stochastic expansions was determined in Ref. (1) using the method developed in Ref. (6). The starting point of the analysis is be the stochastic version of Eq. (1) which we use in the main text and has the following form:

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} + r(n)n + \sqrt{\gamma_n(n)n}\,\eta(t,x). \quad [20]$$

Compared to Eq. (1) we have included a demographic noise term whose strength is determined by $\gamma_n(n)$. The precise dependence of the noise strength on n depends on the details of the birth-death process at the front (see Ref. (1), SI, Sec. IV for an extensive discussion). For example, the simulations presented in this study correspond to $\gamma_n(n) = \gamma_n^{(0)}(1 - n/N)$, but other choices are possible. In general, the stochastic properties of the front are primarily determined by the low density region at the edge of the expansion ($n \ll N$) as we show below, so we expect choice of $\gamma_n(n)$ to not change any of our results as long as $\gamma_n(n) \to \gamma_n^{(0)}$ smoothly as $n \to 0$, for some constant $\gamma_n^{(0)}$.

Solving Eq. (20) directly is difficult because the moments of $n(t,x)$ do not close due to the nonlinear growth term $r(n)$. Instead Ref. (6) argued that the properties of Eq. (20) can be determined using a moment closure scheme, whereby the equation describing the stochastic process $n(t,x)$ is modified to one where the equation for the first moment closes Eq. (1)[§]. It was also shown that the effect of such conditioning can be captured using the equation

$$D\frac{d^2 n}{d\zeta^2} - v\frac{dn}{d\zeta} + r(n)n - \frac{\gamma_n(n)n^2 e^{v\zeta/D}}{\int_{-\infty}^{+\infty} d\xi n^2(\xi) e^{v\xi/D}} = 0. \quad [21]$$

[§]One may wonder why such as scheme would give the correct description of the front dynamics, since we later argue that large fluctuations in the front edge are crucial for recovering the correct behavior of the front velocity and genealogical structure. While this is not at all obvious *a priori*, it is partially justified by Ref. (13) which showed that closing the hierarchy of equations at higher moments of $n(t,x)$ leads to very similar results as the simple first-order closure used in Refs. (1, 6) and here. In addition, this approach was shown to lead to the correct exponents for the scaling of T_c with N in the case of semi-pushed waves (1) and, as we will later show here, also captures the correct structure of the coalescent. These results suggests the moment-closure scheme does give an accurate description of dynamics of Eq. (20), although more work is needed to prove this.

The last term in Eq. (21) is exponentially suppressed for small values of ζ, where the solution is well approximated by the solution to the deterministic equation Eq. (1). At large values of ζ it has the effect of imposing a cutoff on the growth function at ζ_s which can be determined by equating the last two terms:

$$r(n(\zeta_s)) = \frac{\gamma_n(n(\zeta_s))n(\zeta_s)e^{v\zeta_s/D}}{\int_{-\infty}^{+\zeta_s} d\zeta n^2(\zeta)e^{v\zeta/D}}. \quad [22]$$

The above equation can be further simplified by noting that for large values of ζ_s, $n(\zeta_s) \ll N$ and both $r(n)$ and $\gamma_n(n)$ can be approximated by their values at $n = 0$. Making this approximation and using $\rho(\zeta) = n(\zeta)/N$ as before, we obtain the following implicit equation for the cutoff ζ_s:

$$\rho(\zeta_s)e^{v\zeta_s/D} \approx \frac{r_0 N}{\gamma_n^{(0)}} \int_{-\infty}^{\zeta_s} d\zeta \rho^2(\zeta)e^{v\zeta/D}. \quad [23]$$

To solve Eq. (23) note that the integral converges to a constant value with a very weak dependence on ζ_s, which can be ignored to leading order in N. It is also useful to make the replacement $v/D = k + q$ from the definitions in Eq. (12). Substituting this expression along with Eq. (13) into Eq. (23) we get the following expression for the cutoff in pushed stochastic fronts:

$$\zeta_s^{\text{pushed}} \approx \frac{1}{q} \ln N. \quad [24]$$

Note that compared to the cutoff in the deterministic case given by Eq. (19), the cutoff above goes further into the low density region ahead of the front.

The calculation is similar in the pulled case, but the details end up more complicated due to the presence of the sine factor. Again, we use the relation $v_F/D = 2k_F$ and introduce the adimensional lengths $L = k_F\zeta$ to simplify the notation. Then, the expression for the cutoff is given by the implicit equation

$$\frac{L_c}{\pi} \sin\left(\frac{\pi L}{L_c}\right) e^L \approx A' \frac{r_0 L_c^2}{\pi^2 \gamma_n^{(0)}} N \int^{L_c} d\xi \sin^2\left(\frac{\pi\xi}{L_c}\right), \quad [25]$$

where we use the variable L to control the limit $L \to L_c$. Note that we have dropped the lower limit in the integral on the right hand side of Eq. (25) since our expression for $\rho(L)$ from Eq. (17) is only valid when $L \gg 1$. However, due to the exponential factor e^L in the integral from Eq. (23) and the fact that $\lim_{L\to-\infty} \rho(L) = 1$, the contribution to the integral from the lower limit is small compared to that of the upper limit and can be neglected.

While it may not be obvious at first sight, as in the previous subsection, the prefactor in front of e^L on the left hand side of Eq. (25) only contributes a small constant shift in the value of the cutoff L and can be ignored when N is large. Likewise, the integral on the right hand side will be proportional to L_c times some constant C. Using these results, we find the following simplified equation:

$$e^L \approx \frac{CA' r_0}{2\pi^2 k_F \gamma_n^{(0)}} N L_c^3. \quad [26]$$

An approximate solution in the limit of large N for the above equation can be found through the method of dominant balance. Specifically, taking the logarithm on both sides and ignoring the constant ratio on the right we have

$$L \sim \ln N + 3 \ln L_c. \quad [27]$$

We can then look for a solution in the form of a series of successive approximations $L = L_c + L_2 + \ldots$, where each term in the series is much greater than the next. The first order approximation is given by ignoring the second term on the right of Eq. (27), which gives the familiar $L_c \sim \ln N$ we saw in the deterministic case. In order to go beyond the deterministic approximation, we introduce our series expansion into Eq. (27) to get the next order correction

$$L_2 \sim 3 \ln \ln N. \quad [28]$$

Including both leading orders and converting the result back to ζ therefore gives

$$\zeta_s^{\text{pulled}} \sim \frac{1}{k_F} \ln N + \frac{3}{k_F} \ln \ln N. \quad [29]$$

As a check, we indeed see that $\ln N \gg \ln \ln N$ as is required for our approximation to be self-consistent. In principle, the above procedure can be extended to higher orders, with the next level correction being $O(\ln \ln \ln N)$. However, for any biologically relevant values of N, such corrections are much too small to have any impact on our results and are likely to be of the same or lower magnitude as the terms we have ignored in deriving Eq. (29), such as the ratio $\frac{CA' r_0}{2\pi^2 k_F \gamma_n^{(0)}}$ in Eq. (26). But, as we show in the next subsection, not including the second order correction—which is equivalent to using the deterministic front approximation from Eq. (19)—fails to capture the correct statistics of the reproductive values of the front.

C. Effect of deterministic and stochastic cutoffs on the distribution of reproductive values. We are now in a position to analyze how stochasticity in the front shape impacts the distribution of reproductive values in the population. As we show below, the choice of cutoff—determined by the type of front we are considering, as we have showed in the previous subsections—has an important effect on the distribution of reproductive values given by Eq. (9). Instead of analyzing the distribution $p(u)$ directly, it is more convenient to examine the effect of the front cutoff using the definition of $u(\zeta)$. Similar to the previous subsections, the calculations in the pulled regime are slightly more involved and are treated separately. Nevertheless, despite the mathematical differences between the different expansion classes, the most important conclusion of these calculations applies to all types of expansions—namely that deterministic front approximations invariably change the nature of the coalescent at the front when N is large.

We begin by rewritting our definition of $u(\zeta)$ using the normalized population density $\rho(\zeta)$ from the previous sections:

$$u(\zeta) = \frac{\rho(\zeta)e^{v\zeta/D}}{N\int_{-\infty}^{\zeta_c} d\xi \rho^2(\xi)e^{v\xi/D}}. \quad [30]$$

As shown previously, in fully pushed and semi-pushed expansions $\rho(\zeta) \sim e^{-k\zeta}$ when ζ is large, from where we see that $u(\zeta) \sim N^{-1}e^{q\zeta}$. Thus, $u(\zeta)$ is a monotonically increasing function of ζ and has a maximum $u_c \equiv u(\zeta_c)$ corresponding to the location of the cutoff in the population density ζ_c. In the case of stochastic fronts, the cutoff location is given by Eq. (24) and the exponential exactly cancels the N^{-1} factor in the definition of $u(\zeta)$. Therefore, the maximum reproductive value u_c becomes a constant independent of N:

$$u(\zeta_s^{\text{pushed}}) \lesssim O(1). \quad [31]$$

The exact value of the constant $u(\zeta_s^{\text{pushed}})$ in general depends on the details of the model and we expect it to be non-universal. In the case of fully pushed expansions, $p(u)$ has a finite variance and the structure of the coalescent does not depend on the tails of the distribution (8, Sec. 3.2). For semi-pushed expansions, the tail of the distribution $p(u)$ is important so here we might expect the cutoff to have a greater impact. However, in general only the distribution at very large values near the maximum at $u = 1$ will be affected. Such a cutoff is likely to have an impact on the statistics of very rare events when the number of descendants from one individual over a timescale comparable to the mixing time is close to the entire population of the front. But in most cases such events are so rare that they do not significantly change the overall shape of the SFS or the allele frequency distributions we are interested in.

In the case of pushed deterministic expansions however, the situation is very different. Substitution of ζ_d^{pushed} from Eq. (14) into Eq. (30) gives

$$u(\zeta_d^{\text{pushed}}) \sim N^{-\alpha/(1+\alpha)}, \quad [32]$$

where α is the exponent of the distribution of reproductive values defined in Eq. (3) in the main text and Eq. (9). Compared to the stochastic case, now the maximum reproductive value decreases with N. For fully pushed expansions this could change the convergence of the model to the Kingman coalescent, but would otherwise leave structure intact. For semi-pushed expansions, however, the maximum fraction of lineages that can coalesce within a generation goes to zero as $N \to \infty$. In other words, large multiple mergers become increasingly rare in both the fully pushed and semi-pushed deterministic fronts, contrary to the stochastic case.

Finally, we now show that the change in coalescent structure due to a deterministic front approximation we demonstrated for semi-pushed expansions also applies in the pulled regime. The argument is essentially the same as the one presented above, but the expressions we obtain for the different quantities are slightly more complicated due to the presence of the sine factor in front shape in Eq. (15). Readers not interested in the mathematical details of the calculation can simply continue to the next section.

It is again simplest to work using the adimensional lengths. Substituting Eq. (17) into Eq. (30) we get

$$u(\zeta) \approx \frac{\frac{L_c}{\pi}\sin(\pi L/L_c)e^L}{NA'\left(\frac{L_c}{\pi}\right)^2\int_{-\infty}^{L_c} d\xi \sin^2\left(\frac{\pi\xi}{L_c}\right)}. \quad [33]$$

As before, the prefactor in front of e^L can be ignored for L close to L_c and large N, giving the following asymptotic relation

$$u(L) \sim \frac{e^L}{NL_c^3}. \quad [34]$$

The cutoff at $L = L_c$ again imposes a cutoff on the maximum value of u. To determine the value of this cutoff in both stochastic and deterministic fronts, we simply introduce the appropriate value of L_c into Eq. (34). Thus, for stochastic fronts, from Eq. (27) we have $L_s^{\text{pulled}} \sim \ln N + 3\ln\ln N$ which gives the following scaling for the maximum u:

$$u(L_s^{\text{pulled}}) \sim \frac{N\ln^3 N}{N(\ln N + 3\ln\ln N)^3} \lesssim O(1). \quad [35]$$

Similarly to the semi-pushed regime, we find that the cutoff in the stochastic regime only has an effect at very large values of u. To get the corresponding behavior for deterministic fronts, we insert Eq. (19) into Eq. (34). In this case the result reads

$$u(L_d^{\rm pulled}) \sim \frac{N}{N\ln^3 N} \sim \ln^{-3} N. \quad [36]$$

Thus, we see that as in the case of semi-pushed expansions, restricting fluctuations at the front prevents the emergence of highly fecund clones which are crucial for generating non-Kingman genealogies. Unlike in the semi-pushed case, the cutoff at high values of u for pulled deterministic expansions is only logarithmic in N.

As shown by Eqs. Eq. (32) and Eq. (36), the maximum reproductive value in both semi-pushed and pulled expansions goes to zero as $N \to \infty$. However, we are not aware of any rigorous proof that the genealogies in such a model converge to the Kingman coalescent. Indeed, if we naively convert the reproductive values into actual offspring using $W = Nu$, then the variance of W scales as $N^{(1-\alpha)/(1+\alpha)}$ for semi-pushed and as $N/\ln^3 N$ for pulled expansions, so we cannot use the standard results for the convergence to the Kingman coalescent of populations with finite variance (4). Our simulations, however, strongly suggest the emerge of a Kingman-like coalescent, at least for clones with frequencies high enough to sample $p(u)$ near u_c.

2. Effective clone size distribution

In this section we calculate the clone size distribution at the front on times scales much longer than τ_m, by approximating the process in the effective well-mixed population by a branching process. The probability distribution can be obtained analytically for $\alpha = \frac{1}{2}$ and when the number of descendants has a finite variance. All of the results in this section have been derived previously. Some useful references for readers unfamiliar with the theory of branching processes are summarized in A. Readers familiar with this literature may wish to skip parts or all of this section. To help guide readers through the most important results derived in this section we provide a brief overview below.

The definitions of the different types of branching processes and the derivation of the master equation for the general case are discussed in B. In D, we consider a branching process with a finite variance in the number of offspring, which is directly relevant for the Kingman coalescent. Both the survival probability $S(t)$ and the size distribution of surviving clones $(p_+(t, y))$ are introduced and discussed here. In E we give a more formal definition of $p_+(t, y)$ and show how to calculate the complementary cumulative distribution function (CCDF) shown in S4. All of the previous results are extended to processes with infinite variance in the offspring number in F and G. We then use these results to derive the allele frequency distribution for different types of branching processes in 2.

A. Relevant literature. Branching processes were first studied by Watson and Galton to describe the dynamics of British surnames (14); therefore they are often referred to as Galton-Watson processes. Branching processes have been applied to a number of fields including branching of neutrons in nuclear reactions, population genetics, earthquakes, chemical reaction, birth-death processes, shot noise, and many others. The monograph by Harris (15) contains the historical details and detailed mathematical treatment of simple and generalized branching processes together with several applications. A simpler and more limited exposition can be found in Ref. (16). A summary of the early progress in branching processes can be found in Ref. (17). Branching processes were also called multiplicative processes possibly because of the application to the nuclear reactions; see (18).

The full solution for the branching process was developed by a great number of scientists who calculated different properties under different assumptions. Some of the key results that are relevant for us were obtained in Ref. (19, 20). The approach taken in the latter reference is very close to how a physicist would approach this problem and our discussion closely follows that of Ref. (20). More recently, branching processes have been used in the study of avalanches and total popularity on networks (21, 22). These references extend the classical results to compute the integral of the number of organisms over time for surviving families, i.e. avalanche size. On the mathematical side, branching processes can be studied in the continuum limit, which is known as continuous state branching processes. This description is equivalent to a Levy process with a time change. All of the results, however, can be derived from the discrete number of individuals by taking the continuum limit (23–25).

B. Problem formulation and general solution. We consider a continuous time version of the branching process since it is simpler. The probability to observe n individuals at time t is denoted as $p_n(t)$. Unless specified otherwise, we assume that $p_n(0) = \delta_{n,1}$. The probability to leave k descendants is q_k.

The master equation reads

$$\dot{p_k} = r[-kp_k + \sum_{l=0}^{k} q_l(k-l+1)p_{k-l+1}], \quad [37]$$

where r is the branching rate. Since r only enters the problem through the time scale, we set $r = 1$ in the following.

Note that the transition rates are proportional to the number of individuals since each can reproduce. The "+1" in the last term accounts for the fact that the reproducing individual dies.

The master equation can be solved using generating functions. We denote the generating functions for p_n and q_k by P and Q respectively:

$$P(t, z) = \sum_{n=0}^{\infty} z^n p_n(t), \quad [38]$$

$$Q(z) = \sum_{k=0}^{\infty} z^k q_k. \quad [39]$$

Upon differentiating Eq. (38) with time and using Eq. (37), we obtain

$$\frac{\partial P}{\partial t} = [Q(z) - z]\frac{\partial P}{\partial z}, \quad [40]$$

which can be solved using the method of characteristics. Assuming that we start with one individual, $P(0, z) = z$, and the implicit solution of Eq. (40) reads

$$t = \int_z^{P(t,z)} \frac{ds}{Q(s) - s}. \quad [41]$$

This equation serves as the basis of our analysis in the rest of this summary.

Before proceeding with the analysis, however, we point out that many references study branching processes from a different starting point. Consider how the population can change in a short time dt at the start of the process when there is only one individual (similar to the backward Kolmogorov equation). With probability $1 - dt$, nothing happens and the generating function remains unchanged. With probability dt the organism reproduces and leaves k descendants with probability q_k. After that we also have a branching process that lasts time t, but starts with k individuals. Since individuals are independent the generating function for the sum of their progenies is the product of the generating functions for each starting organism. In other words, we obtain

$$P(t + dt, z) = (1 - dt)P(t, z) + dt \sum_{k=0}^{\infty} q_k P(t, z)^k, \quad [42]$$

which simplifies to

$$\frac{\partial P(t, z)}{\partial t} = Q[P(t, z)] - P(t, z). \quad [43]$$

It is easy to see by direct substitution that the implicit solution from Eq. (41) satisfies Eq. (43). The direct analysis of Eq. (43) and its discrete-time analog involves functional equations and recurrences, which are more cumbersome than the implicit solution obtained above.

C. Asymptotic analysis. When the integral in Eq. (41) can be evaluated one can obtain $P(t, z)$ directly. For a general $Q(z)$, we focus on long time limit. In this limit, $t \to +\infty$ and the integral must diverge. Therefore, the long time behavior of $P(t, z)$ is controlled by the smallest root z_c of $Q(z_c) = z_c$ and the behavior of $Q(z)$ around z_c (15, 20).

It is easy to show that $z_c > 1$ when the mean number of descendants $\langle k\rangle = Q'(1) < 1$. In this case, $P(t, z)$ approaches 1 exponentially fast, which corresponds to guaranteed extinction. Note that any generating function needs to be less or equal to one for $|z| \leq 1$.

When $\langle k\rangle = Q'(1) > 1$, $z_c < 1$. In this case, the process has a finite probability to survive, which is given by $1 - z_c$. The population size of surviving realizations grows exponentially with time at a rate given by $\langle k\rangle - 1$. More refined results can be obtained by expanding $Q(z)$ in Taylor series around z_c.

When $\langle k\rangle = Q'(1) = 1$, we have a critical branching process. This is the case that we will focus on in the following. In this case $z_c = 1$ and the behavior of $P(t, z)$ depends on the behavior of $Q(z)$ around $z = 1$. If $\langle k^2\rangle$ exists, $Q(z)$ has a second derivative at $z = 1$ and can be approximated by $Q(z) = z + [1/2Q''(1)](1 - z)^2$. If the variance is infinite, then $Q(z)$ is not analytic around $z = 1$. We argue below that, when the number of descendants is distributed according to a power law, $Q(z) = z + g(1 - z)^{1+\alpha}$ with $\alpha \in (0, 1]$.

In the next two sections, we evaluate the integral in Eq. (41) using the approximations for $Q(z)$ to obtain the long time asymptotics of $P(t, z)$.

D. Critical branching process with finite variance. Upon substituting $Q(z) = z + \frac{Q''(1)}{2}(1 - z)^2$ into Eq. (41) and evaluating the integral, we obtain

$$P(t, z) = 1 - \frac{1 - z}{1 + \frac{Q''(1)t}{2}(1 - z)}. \quad [44]$$

The survival probability is given by

$$S(t) = 1 - P(t, 0) = \frac{1}{1 + Q''(1)t/2} \sim \frac{2}{Q''(1)t}. \quad [45]$$

The average size of a surviving clone $\langle n(t)\rangle_+$ should be such that $\langle n(t)\rangle = 1$. Therefore

$$\langle n(t)\rangle_+ = \frac{1}{S(t)} = 1 + \frac{Q''(1)t}{2}. \quad [46]$$

To obtain $p_n(t)$, we expand $P(t,z)$ in Taylor series around $z=0$. The result for $n>0$ reads

$$p_n(t) = S^{n+1}(t) \sim \left(\frac{2}{Q''(1)t}\right)^2 e^{-\frac{2n}{Q''(1)t}}. \quad [47]$$

The above expression can be recast in a simpler form by normalizing the population size by the expected population size of surviving realizations. Specifically, we let $y = n(t)/\langle n(t)\rangle_+$, which also affects the normalization constant, and divide p_n by $S(t)$ since we consider only surviving realizations. The distribution of scaled population sizes is then described by the following probability density function:

$$p_+(y) = e^{-y}, \quad [48]$$

where we use the $p_+(\cdot)$ to denote the probability distribution function (PDF) conditioned on non-extinction and have omitted the time variable since the equation corresponds to the limit $t \to +\infty$. This relationship can also be derived in a more formal and general way that we describe below.

E. Continuum limit from generating function. As shown in the previous subsection, when the offspring distribution has a finite variance, we can obtain a simple expression for the clone size at $t \to \infty$ by treating $y = n/\langle n(t)\rangle_+$ as a continuous variable. In this subsection we give a more precise definition for $p_+(y)$ from Eq. (48) and show how to calculate the distribution function of y from the generating function $P(t,z)$ defined in Eq. (38).

We begin by formally defining $p_+(t,y)$ as the probability density function of $y(t) = n/\langle n(t)\rangle_+$ at time t, conditioned on non-extinction. First, notice that

$$\Pr(n(t) \le M | n(t) > 0) = \frac{1}{S(t)} \sum_{k=1}^{M} p_k(t) \approx \frac{1}{S(t)} \int_1^M p_k(t) dk \approx \int_1^{M/\langle n(t)\rangle_+} p_+(t,y) dy. \quad [49]$$

Therefore

$$p_+(t,y)dy \approx \frac{p_k(t)}{S(t)}. \quad [50]$$

Note that at long times $\langle n(t)\rangle_+ \gg 1$, and thus $dy = 1/\langle n(t)\rangle_+ \ll 1$, justifying our treatment of y as a continuous variable. Substituting the previous expression into Eq. (50) and using the definition of $\langle n(t)\rangle_+$ from Eq. (46) we obtain the following expression for the PDF of $y(t)$:

$$p_+(t,y) \approx \frac{p_n(t)}{S^2(t)}. \quad [51]$$

Then, we can relate the generating function $P(t,z)$ to the moment generating function of $p_+(t,y)$:

$$\begin{aligned} M(t,\sigma) = \mathrm{E}\{e^{-\sigma y}\} &= \int_0^{+\infty} p_+(t,y) e^{-\sigma y} dy \approx \sum_{n=1}^{+\infty} \frac{p_n(t)}{S(t)} e^{-\sigma n/\langle n(t)\rangle_+} \\ &= \frac{1}{S(t)} \left[P(t, z = e^{-\sigma/\langle n(t)\rangle_+}) - p_0(t)\right] = 1 - \frac{1 - P(t, z = e^{-\sigma/\langle n(t)\rangle_+})}{S(t)}, \end{aligned} \quad [52]$$

where we used $p_0(t) = 1 - S(t)$. One can then obtain $p_+(t,y)$ via an inverse Laplace transform of $M(t,\sigma)$. Note that for the critical branching process $\langle n(t)\rangle_+ = 1/S(t)$.

Since it is convenient to summarize simulation results in terms of the complementary (reverse) cumulative distribution $c(t,y)$, we also derive the connection between $P(t,z)$ and the Laplace transform of $c(t,y)$:

$$C(t,\sigma) = \frac{1 - M(t,\sigma)}{\sigma} = \frac{1 - P(t, z = e^{-\sigma/\langle n(t)\rangle_+})}{S(t)\sigma}. \quad [53]$$

As a check, we can apply this result to the branching process with finite variance to obtain the long time limit of $c(t,y)$ as follows:

$$C(\sigma) = \lim_{t\to+\infty} C(t,\sigma) = \frac{1}{1+\sigma}, \quad [54]$$

$$c(y) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{\sigma y} C(\sigma) d\sigma = e^{-y}, \quad [55]$$

which indeed describes the CCDF for $p_+(y)$.

F. Power-law tails and the behavior of the generating function. Before repeating the analysis above for distributions of the number of descendants q_k with diverging variance, we briefly discuss the connection between the power law tail of q_k and the singularity of $Q(z)$ at $z = 1$. As a reminder, we focus only on critical branching processes with $\langle k \rangle = 1$ and only on $q_k \sim k^{-2-\alpha}$ for large k. Under these assumptions,

$$Q(z) \approx z + g(1-z)^{1+\alpha} \tag{56}$$

around $z = 1$.

The simplest way to show that Eq. (56) is the generating function for a broad offspring distribution is to perform a binomial expansion of the second term and identify the coefficients with the probabilities q_k. After the expansion we have

$$Q(z) = z + g\sum_{k=0}^{\infty}(-1)^k \binom{1+\alpha}{k} z^k, \tag{57}$$

where $\binom{1+\alpha}{k} = \frac{1}{k!}\prod_{j=0}^{k}(1-j+\alpha)$. Note that for $k \geq 2$ all of the factors in the binomial coefficient are negative. We can simplify the expression by separating the first two factors. For the kth term in the series this gives $\prod_{j=0}^{k}(1-j+\alpha) = (-1)^{k-2}\alpha(1+\alpha)\prod_{j=1}^{k-1}(j-\alpha)$. Introducing this expression into Eq. (57) gives

$$Q(z) = g + [1-g(1+\alpha)]z + g\alpha(1+\alpha)\sum_{k=2}^{\infty}\frac{\Gamma(k-1-\alpha)}{\Gamma(1-\alpha)\Gamma(k+1)}z^k, \tag{58}$$

where we have expressed the products over j in terms of Gamma functions. Note that the "-1" factors cancel out as required for q_k to be positive. From Eq. (58) we can simply read off the values of q_k, which for $k \geq 2$ have the form $q_k = g\alpha(1+\alpha)\frac{\Gamma(k-1-\alpha)}{\Gamma(1-\alpha)\Gamma(k+1)}$. To see the scaling of q_k when k is large, we use Stirling's approximation for the Gamma functions and find the following expression:

$$q_k \approx \frac{g\alpha(1+\alpha)}{\Gamma(1-\alpha)}k^{-2-\alpha}. \tag{59}$$

Another way to derive the relationship is to choose a specific form of q_k. A convenient choice is $q_k = k^{-(2+\alpha)}/\zeta(1+\alpha)$ for $k > 0$ and $q_0 = 1 - \zeta(2+\alpha)/\zeta(1+\alpha)$, where $\zeta(\cdot)$ is the Riemann zeta function. Note that this choice satisfies both the normalization condition and the requirement that the average number of descendants equals to one. It is easy to show via a Taylor expansion around $z = 0$ that the corresponding generating function is given by

$$Q(z) = 1 - \frac{\zeta(2+\alpha)}{\zeta(1+\alpha)} + \frac{z}{\zeta(1+\alpha)\Gamma(2+\alpha)}\int_0^{+\infty}\frac{e^{-p}p^{1+\alpha}}{1-ze^{-p}}dp, \tag{60}$$

where the last term without the zeta function is known as $\mathrm{Li}_{2+\alpha}(\cdot)$, polylogarithm of order $2+\alpha$. The asymptotics of $Q(z)$ can be directly extracted from this integral representation by extending it in the complex plane, or from the asymptotics of the polylogarithm.

G. Critical branching process with diverging variance. To find $P(t,z)$, we substitute the approximation for $Q(z)$ (Eq. (56)) into the implicit solution given by Eq. (41). The result reads

$$P(t,z) = 1 - \frac{1-z}{(1+\alpha g t(1-z)^{\alpha})^{1/\alpha}}. \tag{61}$$

This expression contains all the information that we need. In particular, one can pass to a continuum limit and obtain $C(t,\sigma)$ and $c(t,y)$. Inverse Laplace transform can be evaluated by moving the integration contour to hug the branch cut $(-\infty, 0)$. Below, we consider a few special cases where the calculations are particularly simple and provide additional insight.

The survival probability and the average size of the surviving population are given by

$$S(t) = \frac{1}{\langle n(t)\rangle_+} = (1+\alpha g t)^{-1/\alpha} \sim t^{-1/\alpha}. \tag{62}$$

Note that the relevant time scale is $1/(\alpha g)$, which becomes $\zeta(1+\alpha)(1+\alpha)/\Gamma(1-\alpha)$. The latter expression scales as $1/\alpha$ for $\alpha \to 0$. Thus, one should expect very long transient dynamics for small α.

The long time limit for $C(t,\sigma)$ is given by

$$C(\sigma) = \frac{1}{(1+\sigma^{\alpha})^{1/\alpha}}. \tag{63}$$

The inverse Laplace transform yields the following asymptotics

$$c(y) \sim \begin{cases} 1 - \dfrac{y^\alpha}{\alpha\Gamma(1+\alpha)}, & y \ll 1, \\ \dfrac{y^{-1-\alpha}}{\Gamma(1-\alpha)}, & y \gg 1. \end{cases} \quad [64]$$

The asymptotics for $p_+(y)$ are obtained by differentiation with respect to y.

For the special case of $\alpha = 1/2$, one can obtain an analytic expression for $c(y)$:

$$c(y) = (1+2y)e^y \operatorname{erfc}(\sqrt{y}) - 2\sqrt{\frac{y}{\pi}} \sim \begin{cases} 1 - \dfrac{4}{\sqrt{\pi}}\sqrt{y}, & y \ll 1, \\ \dfrac{y^{-3/2}}{\sqrt{\pi}}, & y \gg 1. \end{cases} \quad [65]$$

The small y asymptotics can also be derived directy from the generating function by expanding it in Taylor series around $z = 0$. This yields

$$p_n(t) = (\alpha g t)^{-1-1/\alpha} \frac{\Gamma(n+\alpha)}{\Gamma(1+\alpha)\Gamma(n+1)} \sim (\alpha g t)^{-1-1/\alpha} n^{-1+\alpha}. \quad [66]$$

H. Allele frequency distributions. The results from the previous subsection allow us to derive the properties of the allele frequency distribution (AFD) shown in Fig. 5 in the main text. Below we derive the AFD for semi-pushed and fully pushed waves ($\alpha > 0$). Due to the slow relaxation to the quasi-stationary regime shown in the previous section, the pulled case (corresponding to $\alpha = 0$) is more difficult to analyze in this framework. Interested readers are invited to consult Refs. (26) and (5) for more details on this regime.

To calculate the AFD for the two-allele model we used in the main text, we approximate the sizes of the two alleles by independent branching processes $n_1(t)$ and $n_2(t)$, with identical distributions of reproductive values given by Eq. (2) in the main text. Note that in the full model the total population size of the front is kept constant, which induces correlations between the number of offspring of the two alleles. But our solution should still capture the shape of the distribution close to the edges of frequency range, where the size of one allele is small and the correlations weak. Using this approximation, we define the AFD $A(z)$ as the PDF of the ratio $Z_t = \frac{n_1(t)}{n_1(t)+n_2(t)}$ conditioned on both $n_1(t)$ and $n_2(t)$ surviving. The distribution can be expressed in terms of the complementary cumulative distributions derived earlier as follows:

$$\Pr\{Z_t > z\} = \Pr\left\{\frac{n_1(t)}{n_1(t)+n_2(t)} > z\right\} = \Pr\left\{n_1(t) > \frac{z}{1-z} n_2(t)\right\} = \int_0^\infty dy\, p_+(y) \int_{\frac{zy}{1-z}}^\infty dx\, p_+(x), \quad [67]$$

where we defined $x = n_1(t)/\langle n_1(t)\rangle_+$ and $y = n_2(t)/\langle n_2(t)\rangle_+$ and used $\langle n_1(t)\rangle_+ = \langle n_2(t)\rangle_+ = 1/S(t)$ to express the probability in terms of the PDF defined in Eq. (51). Using the definition of the CCDF from Eq. (55) we find the following expression for the allele frequency distribution

$$\Pr\{Z_t > z\} = \int_0^\infty dy\, p_+(y) c\left(\frac{z}{1-z} y\right). \quad [68]$$

We first consider the case where the variance is finite, when both $p_+(y)$ and $c(y)$ are exponential. The calculation is straightforward and gives

$$\Pr\{Z_t > z\} = \int_0^\infty dy\, \exp\left(-y - \frac{z}{1-z} y\right) = 1 - z. \quad [69]$$

Thus, for branching processes with finite variance Z_t is uniformly distributed on the interval $[0, 1]$.

For the general branching process with infinite variance Eq. (68) cannot be evaluated exactly since there are no closed form expressions for the clone size distribution. Nevertheless, we can use the asymptotic expressions from Eq. (64) to obtain the scaling of the allele frequency distribution close to $z = 0$ and $z = 1$. We begin by differentiating Eq. (64) to find the asymptotics for $p_+(y)$. The result reads

$$p_+(y) \sim \begin{cases} \dfrac{y^{\alpha-1}}{\Gamma(1+\alpha)}, & y \ll 1, \\ \dfrac{1+\alpha}{\Gamma(1-\alpha)} y^{-2-\alpha}, & y \gg 1. \end{cases} \quad [70]$$

We now show that for z very small or close to 1, the asymptotic scalings from Eq. (64) and Eq. (70) are enough to find the leading behavior. Note that since $n_1(t)$ and $n_2(t)$ have identical distributions, the distribution of z will be symmetric about $z = 1/2$. Thus, we focus on $z \ll 1$ in the following for simplicity. To simplify the calculations we use the notation $u(z) = z/(1-z)$, which we will substitute back at the end to obtain the distribution in terms of z. Denoting the integrand by

$f(y,u) = p_+(y)c(uy)$, there are three regions of y where we can use the asymptotic relations derived previously to find the dependence of f on y and u:

$$f(y,u) \sim \begin{cases} \dfrac{y^{\alpha-1}}{\Gamma(1+\alpha)} - \dfrac{u^\alpha}{\alpha\Gamma^2(1+\alpha)} y^{2\alpha-1}, & y \ll 1, \\ \dfrac{1+\alpha}{\Gamma(1-\alpha)} y^{-2-\alpha} - \dfrac{(1+\alpha)u^\alpha}{\alpha\Gamma(1-\alpha)\Gamma(1+\alpha)} y^{-2}, & 1 \ll y \ll u^{-1}, \\ \dfrac{(1+\alpha)u^{-1-\alpha}}{\Gamma^2(1-\alpha)} y^{-3-2\alpha}, & uy \gg 1. \end{cases} \quad [71]$$

From the above expression we immediately see that the leading order for small u comes from the two regions where $y \ll u^{-1}$ and scales as u^α. Substituting z from the definition of u and noting that $\Pr\{Z_t > 0\} = 1$ by definition, we find the following asymptotic scaling for the distribution of Z_t:

$$\Pr\{Z_t > z\} \sim 1 - C_\alpha z^\alpha. \quad [72]$$

The constant C_α will in general depend on the crossover between three regimes in Eq. (71) and acts as a fitting parameter in our simulations. A similar calculation in the limit of $u \gg 1$ confirms that $\Pr\{Z_t > z\} \sim (1-z)^\alpha$ for $1 - z \ll 1$, as expected from the symmetry argument. Differentiating with respect to z then gives the expression for the allele frequency distribution near the boundaries:

$$A(z) \sim \begin{cases} \alpha C_\alpha z^{\alpha-1}, & z \ll 1, \\ \alpha C_\alpha (1-z)^{\alpha-1}, & 1 - z \ll 1. \end{cases} \quad [73]$$

For fitting Eq. (73) to the simulations in Figs. 5 from the main text and S3 we used $A(z) = Cz^\alpha(1-z)^\alpha$ and set C such that the integral $\int_{z_0}^{1-z_0} dz A(z) = 1$. This then leaves z_0 as a fitting parameter, which we chose such that the distribution approximately matched the height of the bins at the edges of the histograms in the figures.

3. Summary statistics of ancestral trees

In this section we present other summary statistics we used to infer the topology of the ancestral trees obtained from simulations.

A. Theoretical background. Our analysis of the genealogies is based on the coalescent theory. The coalescent provides a model for the backward-in-time dynamics of lineages in a population without any internal structure [¶]. Generally, such a model is completely described by the rates $\lambda_{b,k}$ at which k out of b lineages merge. An important result shows that $\lambda_{b,k}$ for any coalescent[‖] can be written in the following form:

$$\lambda_{b,k} = \int_0^1 dx x^k (1-x)^{b-k} \frac{\Lambda(x)}{x^2}, \quad [74]$$

where $\Lambda(x)/x^2$ is the distribution of the merger sizes (8). A few special choices of $\Lambda(x)$ are worth noting. First, $\Lambda(x) = \delta(x)$ gives $\lambda_{b,2} = 1$ and $\lambda_{b,k} = 0$ for $k > 2$. This is the standard Kingman coalescent, where only pairwise mergers are allowed and their rate is constant. Another important model is the Bolthausen–Sznitman coalescent and is given by $\Lambda(x) = 1$. The merger rates in this case are $\lambda_{b,k} = \frac{(b-1)!}{(k-2)!(b-k)!}$, which implies that $k = 2$ and $k = b$ mergers are equally likely and the most likely merger size is close to $\frac{b}{2}$. Such large merger events have been used to describe genealogies of populations under strong selection (26, 32). Finally, one can interpolate between the two by using

$$\Lambda(x) = \frac{x^{1-\beta}(1-x)^{\beta-1}}{\Gamma(2-\beta)\Gamma(\beta)}. \quad [75]$$

From Eq. (75), it is easy to show that $\beta = 1$ gives the Bolthausen–Sznitman coalescent and $\beta = 2$ gives the Kingman coalescent. The coalescent described by Eq. (75) is known as the Beta-coalescent and many of its properties have been studied previously (8). For our purposes, it is important to note that the Beta-coalescent describes the genealogies of highly fecund populations, in which the offspring distribution $P(W)$ (which is equivalent to our distribution of reproductive values $P(u)$) has a power law tail $P(W) \sim W^{-(1+\beta)}$ (4). In Sec. 1, we demonstrated that the distribution of reproductive values indeed has such a power law tail in range expansions, when coarse-grained over the mixing time. We make use of this fact to argue that genealogies in range expansions can generically be described by a Beta-coalescent.

[¶] Mathematically, this property is referred to as exchangeability and is an underlying premise in coalescent theory.

[‖] This result applies to all Λ-coalescents, in which any number of lineages can merge at the same time, but merger events happen in succession. An even more general class, known as the Ξ-coalescent, allows for multiple simultaneous merger events as well. Such models have mainly been used to describe genealogies of diploid populations (27–29), but also populations under strong selection in the presence of recombination (30, 31).

B. One- and two-site frequency spectra. The most convenient way to characterize the statistical properties of genealogical trees is the site frequency spectrum (SFS), which corresponds to the set of lengths of branches ξ_k subtending k leaves, for all values of $k \in [1, n-1]$. The exact SFS can be obtained recursively for small n (33). Asymptotic results for large n are also known for Eq. (75), but they converge slowly with sample size (33) and we will not use them here. Similar results can be obtained for 2-SFS, which represents the covariances between branch lengths (33).

While the SFS of the Beta- and Kingman coalescents are quite different as we have shown, relaxing the assumption of constant population size in the Kingman coalescent can lead to the site frequency spectra becoming more similar. Recently, it has been proposed that the two-site frequency spectrum (2-SFS) is a more robust measure to distinguish between Kingman and non-Kingman coalescents (34). Empirically, one can estimate the 2-SFS $p_n(k,l)$ from a sample of size n sequences by counting the number of pairs of sites which have allele counts k and l. For constant mutation rates, the 2-SFS can be derived from the genealogical tree—in this case $p_n(k,l)$ is proportional to the second moment of the length of branches that subtend k and l leaves. In the case of the Kingman coalescent, the long branches near the common ancestor lead to a large number of sites which co-occur or split the tree in half. This explains the high values of the 2-SFS seen on the diagonals. In addition, pairwise branching of ancestral lineages constrain the topology further down tree, leading to anticorrelations between rare alleles (Fig. S1a). In contrast, coalescents with multiple mergers have shorter branches near the common ancestor, decreasing the density along the diagonal of the 2-SFS. The tree topology of coalescents with multiple mergers is also less constrained by early mergers, resulting in less pronounced negative correlations between rare alleles in the Beta-coalescent, and positive correlations in the Bolthausen–Sznitman coalescent (Fig. S1b, c).

We used the trees generated from simulations of fully pushed, semi-pushed, and pulled expansions to test 2-SFS against the theoretical predictions. We found that the patterns in the 2-SFS qualitatively matched the theoretical predictions for Kingman, Beta-, and Bolthausen–Sznitman coalescents (Fig. S1). In particular, fully pushed waves showed negative correlations outside of the main diagonals as expected, with correlations below the main diagonal smaller in absolute value than those above the main diagonal. Semi-pushed and pulled expansions, on the other hand, showed signatures of multiple mergers in the form of an increase of correlations below the main diagonal, and higher positive correlations on the off-diagonal, especially in the case of pulled expansions. We have attempted to capture these patterns of variation using different summary statistics presented in Fig. S2 and found that the change in these statistics as the coalescent transition from the Kingman to the Bolthausen–Sznitman coalescent in simulations qualitatively agrees with the theory.

4. Simulations

In this section we explain the details of our expansion simulations and the data collection and processing pipelines.

A. Stochastic front simulations. We simulated the expansion of a population in a one-dimensional habitat modeled by an array of patches (demes), separated by a distance Δx. Demes contain individuals, which are labeled using integers. We denote by $I_i(t,x)$ the label of the ith individual in deme x, with $1 \le x \le L$ and $1 \le i \le N$. To allow for demes with less than N individuals, we use vacancies, which are labeled by $I^v = 0$.

The population is initially localized on $L/2 = 150$ demes. Each deme is filled with N individually labeled members of the population. Individuals are labeled sequentially, starting with the first individual in the leftmost deme and moving to the right of the population. Thus,

$$\begin{aligned} I_i(0,x) &= (x-1)N + i + 1, \quad && x \le L/2 \\ I_i(0,x) &= 0, && x > L/2. \end{aligned} \qquad [76]$$

Each generation is updated in two steps. First, a migration step, in which demes are updated sequentially, starting from $x = 1$. For each deme, the number of migrants exchanged with the next deme is drawn from a binomial distribution:

$$n_x^{\text{migrants}} = \text{Binomial}(n_x, m/2), \qquad [77]$$

where m is the migration probability. To choose the migrants, the order of individuals in demes x and $x+1$ is randomized, and the first n_x^{migrants} from the demes are exchanged.

Second, we perform a growth step. Following Ref. (3), the growth of the population was modeled by introducing a fitness difference between the vacancies and the actual species. Specifically, the fitness of the species was set to $w_x^{\text{s}} = 1$ and the fitness of the vacancies was set to $w_x^{\text{v}} = 1 - r(n_x)/(1 - n_x/N)$, where

$$r(n) = r_0(1 - n/N)(1 + Bn/N), \qquad [78]$$

$$n_x(t) = \sum_{j=1}^{N} (1 - \delta_{0, I_j(t,x)}), \qquad [79]$$

and δ_{lm} is the Kronecker delta. The next generation is constructed by sampling, with replacement, a new set of labels $I_i(t+1,x)$ from the set of previous labels $\{I_1(t,x), I_2(t,x), ..., I_N(t,x)\}$ for each $i \le N$. The probability to sample the ancestor $I_i(t,x)$ is proportional to the ratio of w_{ix} to the mean fitness of the population in the deme: $\bar{w}_x = n_x/N + w_x^{\text{v}}(N - n_x)/N = 1 - r(n_x)$.

The parameters for the simulations in each of the three regimes were chosen so as to provide the clearest differences between the coalescents. Specifically, we used purely logistic growth for pulled expansions ($B = 0$ in Eq. 4 in the main text) and a high value of the cooperativity ($B = 10$) for the fully pushed expansions. The cooperativity in the semi-pushed regime ($B = 3.33$) was chosen close to the middle of the semi-pushed regime, where we expected the leading behavior for the coalescence time[**] predicted by theory to be a good approximation to the simulations. As shown in Ref. (1), for B in this region, the predicted exponents for the scaling of T_c with N in both deterministic and stochastic fronts match the theoretical predictions exactly. Note that while our theoretical predictions are valid to leading order for all expansions, close to the transitions between the different classes we expect subleading corrections to play an important role in the dynamics at the front, which can lead to substantial changes in both the coalescence time and the genealogies for the values of N accessible in simulations.

B. Deterministic front simulations. Simulations with deterministic fronts were performed using the same algorithm as in Ref. (1). Briefly, in these simulations, only two genotypes were considered, whose densities are $n_x^{(i)}(t)$, where x is the index of the deme, and $i = 1, 2$ denotes the genotype. The total population for each deme after an update, $n_x(t+1) = n_x^{(1)}(t+1)+n_x^{(2)}(t+1)$, was determined using a similar idea of introducing vacancies in demes of size N as in the stochastic model. Specifically, we calculate the expected population densities following migration

$$\tilde{n}_x^{(i)}\left(t+\frac{1}{2}\right) = (1-m)n_x^{(i)}(t) + \frac{m}{2}n_{x-1}^{(i)}(t) + \frac{m}{2}n_{x+1}^{(i)}(t) \quad [80]$$

and growth

$$\hat{n}_x^{(i)}(t+1) = [1 + r(\tilde{n}_x)]\tilde{n}_x^{(i)}\left(t+\frac{1}{2}\right), \quad [81]$$

where we defined $\tilde{n}_x = \tilde{n}_x^{(1)}(t+1/2) + \tilde{n}_x^{(2)}(t+1/2)$. We then determine the total population density at deme x in the next generation as the closest integer to $\hat{n}_x(t+1) = \hat{n}_x^{(1)}(t+1) + \hat{n}_x^{(2)}(t+1)$. Finally, denoting the updated population density by $n_x(t+1)$, we calculate the frequency of each genotype using binomial sampling:

$$\begin{aligned} n_x^{(1)}(t+1) &= \text{Binom}\left(n_x(t+1), f_x\right), \\ n_x^{(2)}(t+1) &= n_x(t+1) - n_x^{(1)}(t+1), \end{aligned} \quad [82]$$

where

$$f_x = \frac{\hat{n}_x^{(1)}(t+1)}{n_x(t+1)}. \quad [83]$$

C. Recording genealogies. The genealogy of the population is recorded in a custom tree class, in which all individuals in the simulation box are stored as nodes. Each node is assigned a unique parent node, and a set of child nodes, except for the most recent generation, which have no children—we will refer to these nodes as the leaves of the tree. The tree is initialized with one node, which is designated as the root of the tree, and is continuously updated as follows. At the start of the simulation, all individuals at the front are assigned as leaves with the root as their parent. As the population expands, many labels become extinct and the average clone size of the surviving labels grows. After a fixed number of generations Δt, we relabel all individuals and add them as new nodes on the tree. Each individual is assigned as a child node to one of the leaves of the tree, which is designated as its parent according to the previous label of the child node. After every individual is assigned to the tree, the newly added nodes are designated as the new leaves of the tree. At the end of this process, we prune the tree by removing all nodes which have no leaves among their descendants. The process is repeated until either the whole population has one common ancestor or a maximum number of generations $T_{\max}$ for the simulation is reached.

5. Data analysis

In this section we explain how we analyze the data from simulations to obtain the figures in the main text and the SI.

A. Estimating the mixing time. We used the following procedure to determine the spatial distribution of ancestors in Fig. 2 in the main text. We ran 1000 simulations of a fully pushed expansion, for which we estimated the coalescence time $T_c \approx 10^3$, using the following parameters: $N = 350$, $B = 10$, $r_0 = 0.01$, $m = 0.4$, $\Delta t = 20$. For each simulation we recorded the ancestry as described in Sec. 4. In order to determine the location of each ancestor from the population, we modified the label assignment algorithm by using the following equation:

$$I_i(t, x) = NL(t-1) + Nx + i + 1, \quad [84]$$

where $I_i(t, x)$ is the label of individual i from deme x in generation t. Using this equation, each label uniquely specifies the position of the individual.

[**] Note that the coalescence time is defined as $T_c = \Lambda^{-1}$, where Λ is the rate of diversity loss, as defined in Ref. (1).

Because fronts are stochastic it is difficult to compare ancestral distributions across simulations. To minimize the variance in the ancestral distribution due to variations in the final sampling location, we used the following procedure. We first determined the midpoint of the front, given by the deme closest to the mean position x along the front, weighted by the population size at x. Next, we determined the bulk and leading edge of the front, which we defined as least advanced location with population size below the carrying capacity and the most advanced location with a non-zero population size, respectively. Finally, the sampling location from the bulk and the front were chosen as the closest demes to the halfway distance between midpoint of the front, and the bulk edge and the front edge, respectively.

We then collect the labels of all individuals from the two sampling locations. Using the ancestral trees, we traced back the labels of the ancestors of all the sampled individuals. Finally, we recorded the locations of these ancestors by solving for x in Eq. (84) and plotted the distribution of these locations across all simulations. The mixing time was determined approximately as the time at which the two ancestor distributions visually overlapped. During testing, we also tried more quantitative ways of measuring the overlap between the distributions, but ultimately did not find that they were more useful than the simple visual test.

B. Sampling and analysis of SFS and 2-SFS. We used the following procedure to sample and analyze the SFS and 2-SFS from the ancestral trees in our simulations. We first subsampled a number of individuals n from close to the edge of the front in the final population. As discussed in the main text, far from the front the effects of spatial structure become important and our well-mixed approximation breaks down. Empirically, we observed that sampling individuals from the farthest advanced 20 demes minimized the effects of spatial structure on both the SFS and the clone size distributions shown in Fig. S4. The value of n was chosen small enough to allow for comparison with the exact predictions for the different colescent classes described below. Each ancestral tree was sampled independently 10 times in order to obtain better estimates for the averaged quantities we calculated.

The simulations used to generate the ancestral trees were performed choosing three values of B (10, 3.33, and 0, respectively) in Eq. 4 from the main text for each class of waves. Using Eq. 3 from the main text, the values of α for each of these expansions are approximately 1, 0.67, and 0, respectively. The coalescents for well-mixed populations with these descendant distributions are described by the Beta-coalescent from Eq. (75) with the paramter β equal to 2, 1.67, and 1, respectively. To calculate the theoretical predictions for the SFS and 2-SFS we adapted a numerical implementation of the exact recurrence relations for the SFS and 2-SFS from Ref. (34), which was originally developed in Ref. (33).

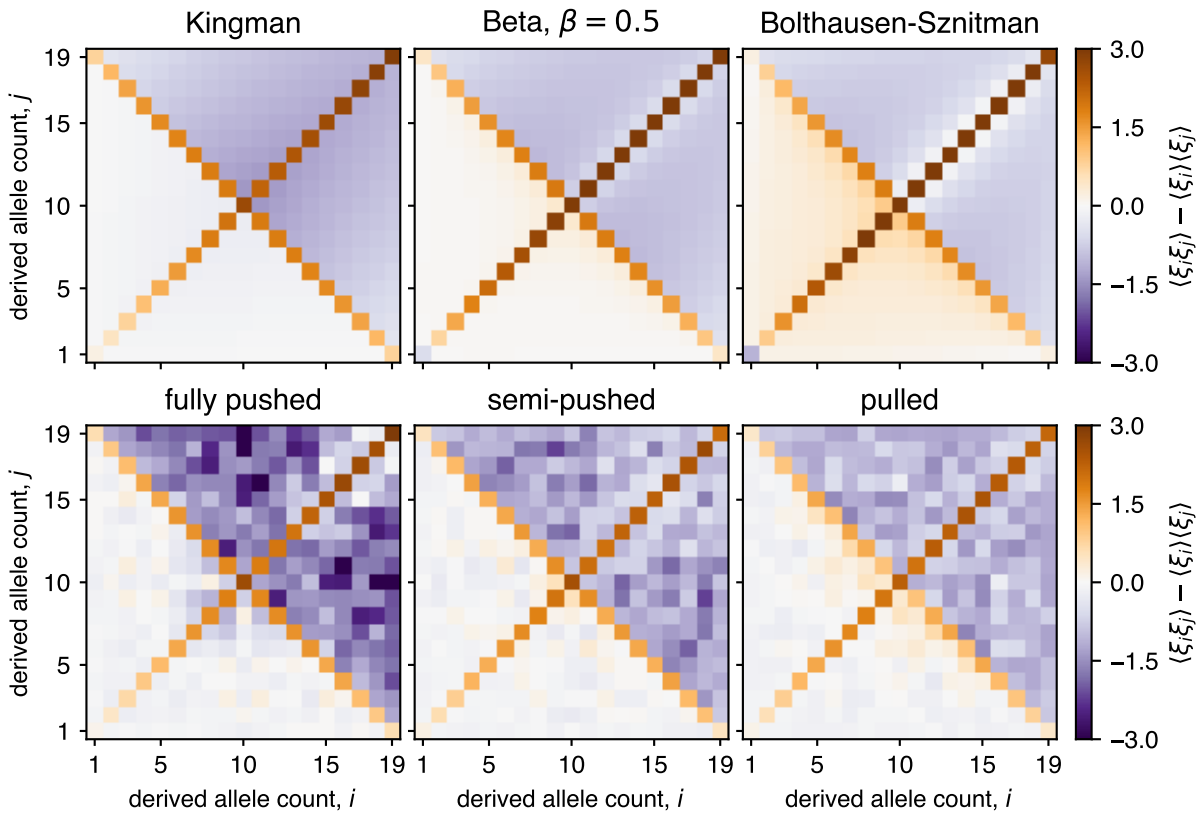

Fig. S1. Comparison of two-site frequency spectra reveal signatures of multiple mergers in semi-pushed and pulled expansions. Matrices show the correlation function between tree branches subtending different number of leaves for both the expected coalescents (top) and expansion simulations (bottom) for each expansion regime. The averaged 2-SFS from simulations were generated using the same sampling procedure used for the SFS (SI, Sec. 4).

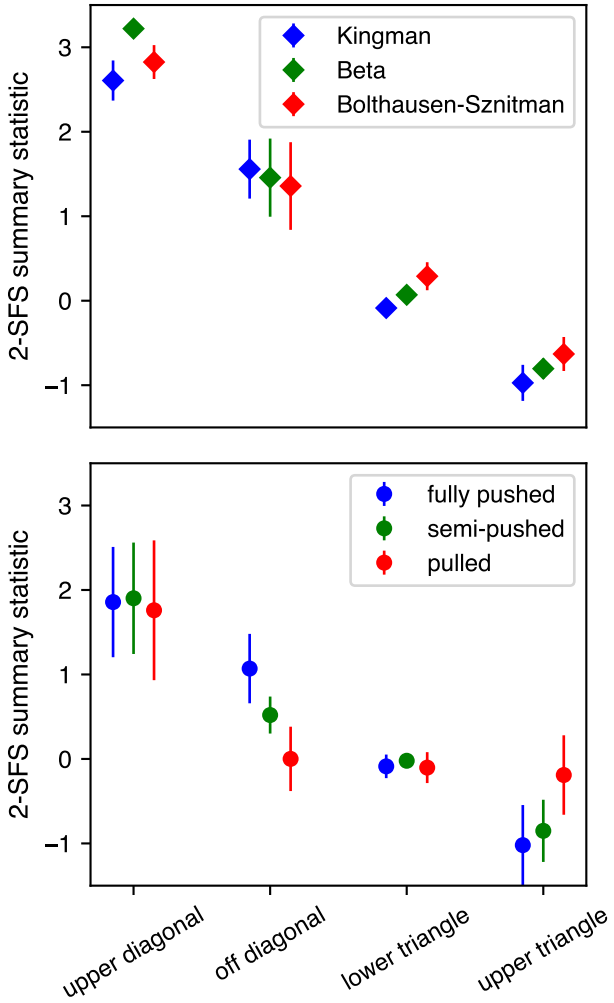

Fig. S2. Summary statistics of 2-SFS show qualitative agreement with theoretical predictions. (Top) Mean values and standard deviation of entries in the 2-SFS for the Kingman (blue), Beta- with $\beta = 0.5$ (green), and Bolthausen–Sznitman (red) coalescents. For all three coalescents a sample size of $n = 20$ was used. The four bins are defined as follows: upper diagonal $= \{(i,i) : \lfloor n/2 \rfloor < i < n\}$, off diagonal $= \{(i, n-i) : 1 \leq i < \lfloor n/2 \rfloor \quad \text{or} \quad \lfloor n/2 \rfloor < i < n\}$, lower triangle $= \{(i,j) : i + j < n - 1, i \neq j\}$, upper triangle $= \{(i,j) : i + j > n - 1, i \neq j\}$. (Bottom) Same as upper panel, but using 2-SFS from simulations of fully pushed (blue), semi-pushed (green) and pulled (red) expansions. All simulation parameters are identical to those for Fig. S1.

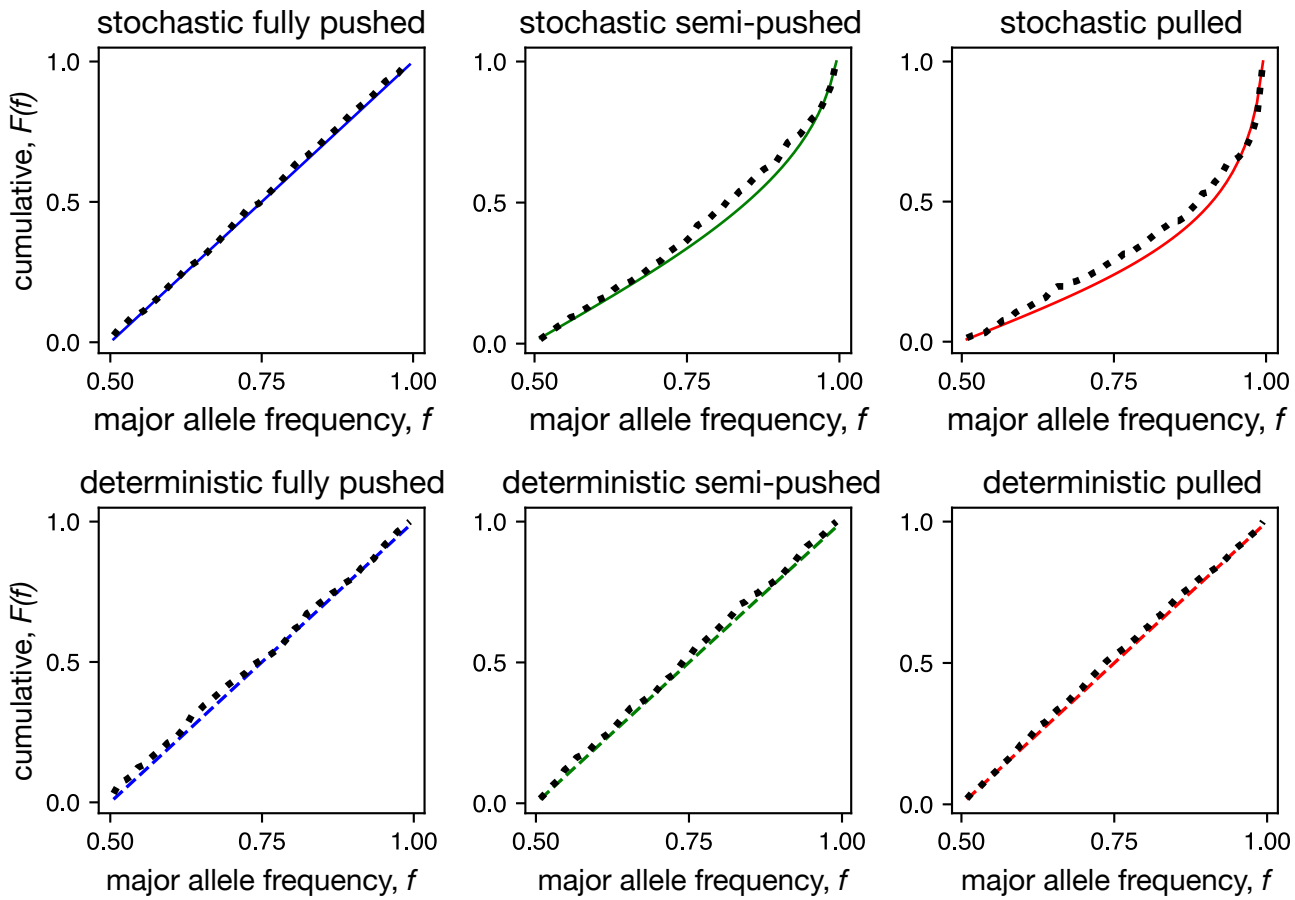

Fig. S3. Allele frequency distributions quantitatively agree with theoretical predictions in both stochastic and deterministic regimes. Shows the same data as Fig. 5 in the main text as a cumulative distribution for better quantitative comparison between theoretical prediction and simulations. Simulations were carried out using the following parameters: $N = 10^6$, $r_0 = 0.01$, $m = 0.4$, $B = 10$ (fully pushed), $B = 3.33$ (semi-pushed), and $B = 0$ (pulled). All simulations were started with equal frequency of the two alleles across the front. Distributions here and in Fig. 5 (main text) are shown after $3,980,000$ (fully pushed), $1,527,315$ (semi-pushed), and $98,827$ (pulled) generations from the start for stochastic waves and after $4,980,000$ (fully pushed), $1,507,719$ (semi-pushed), and $531,529$ (pulled) generations for deterministic waves.

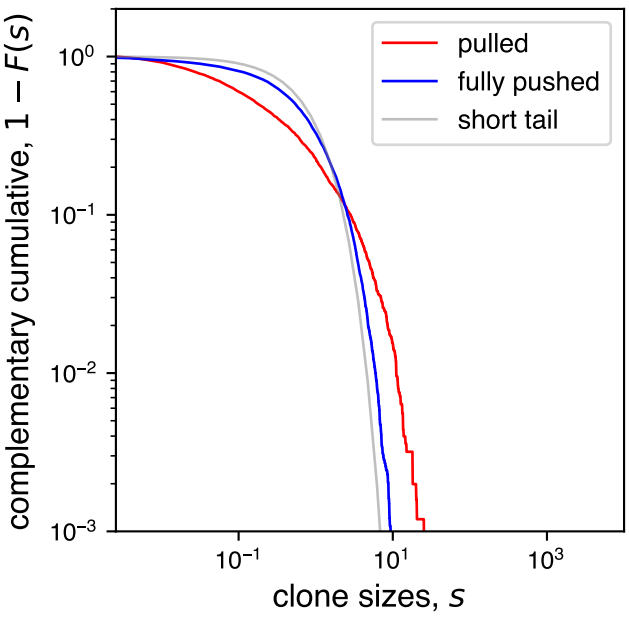

Fig. S4. Pulled expansions have broader clone size distribution compared to fully pushed expansions. The complementary cumulative distribution function of the normalized clone size s (where the normalization is with respect to the mean clone size) for fully pushed and pulled expansions. A total of 100 simulations were run without relabeling individuals and the sizes of distinct clones at the edge of the front were recorded every 500 generations. The front was defined as the first 25 demes starting from the most advanced occupied deme. The growth function used is given by Eq. 4 from the main text, with $B = 10$ (fully pushed) and $B = 0$ (pulled) and all other parameters kept constant. The values of the other simulation parameters were $N = 9600, r_0 = 0.01, m = 0.4$.

References

1. G Birzu, O Hallatschek, KS Korolev, Fluctuations uncover a distinct class of traveling waves. *Proc. Natl. Acad. Sci.* **115**, E3645–E3654 (2018).
2. L Roques, J Garnier, F Hamel, EK Klein, Allee effect promotes diversity in traveling waves of colonization. *Proc. Natl. Acad. Sci.* **109**, 8828–8833 (2012).
3. O Hallatschek, DR Nelson, Gene surfing in expanding populations. *Theor. Popul. Biol.* **73**, 158–170 (2008).
4. J Schweinsberg, Coalescent processes obtained from supercritical galton–watson processes. *Stoch. processes their Appl.* **106**, 107–139 (2003).
5. O Hallatschek, Selection-like biases emerge in population models with recurrent jackpot events. *Genetics* **210**, 1053–1073 (2018).
6. O Hallatschek, The noisy edge of traveling waves. *Proc. Natl. Acad. Sci.* **108**, 1783–1787 (2011).
7. G Birzu, S Matin, O Hallatschek, KS Korolev, Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecol. Lett.* **22**, 1817–1827 (2019).
8. N Berestycki, Recent progress in coalescent theory. *Ensaios Matematicos* **16**, 1–193 (2009).
9. E Brunet, B Derrida, Shift in the velocity of a front due to a cutoff. *Phys. Rev. E* **56**, 2597 (1997).
10. W Van Saarloos, Front propagation into unstable states. *Phys. Reports* **386**, 29–222 (2003).
11. E Brunet, B Derrida, A Mueller, S Munier, Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys. Rev. E* **73**, 056126 (2006).
12. E Brunet, B Derrida, AH Mueller, S Munier, Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76**, 041104 (2007).
13. O Hallatschek, L Geyrhofer, Collective fluctuations in the dynamics of adaptation and other traveling waves. *Genetics* **202**, 1201–1227 (2016).
14. HW Watson, F Galton, On the probability of the extinction of families. *The J. Anthropol. Inst. Gt. Br. Irel.* **4**, 138–144 (1875).
15. TE Harris, *The theory of branching processes.* (Courier Corporation), (2002).
16. R Durrett, *Probability: theory and examples.* (Cambridge university press), (2010).
17. TE Harris, Some mathematical models for branching processes, (RAND CORP SANTA MONICA CA), Technical report (1950).
18. R Otter, The multiplicative process. *The Annals Math. Stat.* **20**, 206–224 (1949).
19. AN Kolmogorov, On the solution of a biological problem. *Proc. Tomsk Univ.* **2**, 7–12 (1938).
20. VM Zolotarev, More exact statements of several theorems in the theory of branching processes. *Theory Probab. & Its Appl.* **2**, 245–253 (1957).
21. KI Goh, DS Lee, B Kahng, D Kim, Sandpile on scale-free networks. *Phys. Rev. Lett.* **91**, 148701 (year?) Publisher: American Physical Society.
22. JP Gleeson, WT Lee, JA Ward, KP O'Sullivan, Competition-induced criticality in a model of meme popularity. *Phys. Rev. Lett.* **112**, 048701 (2014).
23. C Foucart, O Hénard, , et al., Stable continuous-state branching processes with immigration and beta-fleming-viot processes with immigration. *Electron. J. Probab.* **18** (2013).
24. Z Li, Continuous-state branching processes. *arXiv preprint* **2**, 1202.3223 (2012).
25. AE Kyprianou, JC Pardo, Continuous-state branching processes and self-similarity. *J. Appl. Probab.* **45**, 1140–1160 (2008).
26. RA Neher, O Hallatschek, Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci.* **110**, 437–442 (2013).
27. M Möhle, S Sagitov, Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47**, 337–352 (year?).
28. M Birkner, J Blath, B Eldon, An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics* **193**, 255 (year?).
29. M Birkner, H Liu, A Sturm, , et al., Coalescent results for diploid exchangeable population models. *Electron. J. Probab.* **23** (2018).
30. R Durrett, J Schweinsberg, Approximating selective sweeps. *Theor. Popul. Biol.* **66**, 129–138 (year?).
31. R Durrett, J Schweinsberg, A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stoch. Process. their Appl.* **115**, 1628–1657 (year?).
32. MM Desai, AM Walczak, DS Fisher, Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* **193**, 565–585 (2013).
33. M Birkner, J Blath, B Eldon, Statistical Properties of the Site-Frequency Spectrum Associated with lambda-Coalescents. *Genetics* **195**, 1037 (2013).
34. DP Rice, J Novembre, MM Desai, Distinguishing multiple-merger from Kingman coalescence using two-site frequency spectra. *bioRxiv preprint* **11**, 461517 (2018).