Dynamic sampling bias and overdispersion induced by skewed offspring distributions

Takashi Okada[1,2] and Oskar Hallatschek[1]

[1] *Departments of Physics and Integrative Biology, University of California, Berkeley, California 94720*

[2] *RIKEN iTHEMS, Wako, Saitama 351-0198, Japan*

Natural populations often show enhanced genetic drift consistent with a strong skew in their offspring number distribution. The skew arises because the variability of family sizes is either inherently strong or amplified by population expansions, leading to so-called 'jackpot' events. The resulting allele frequency fluctuations are large and, therefore, challenge standard models of population genetics, which assume sufficiently narrow offspring distributions. While the neutral dynamics backward in time can be readily analyzed using coalescent approaches, we still know little about the effect of broad offspring distributions on the dynamics forward in time, especially with selection. Here, we employ an exact asymptotic analysis combined with a scaling hypothesis to demonstrate that over-dispersed frequency trajectories emerge from the competition of conventional forces, such as selection or mutations, with an emerging time-dependent sampling bias against the minor allele. The sampling bias arises from the characteristic time-dependence of the largest sampled family size within each allelic type. Using this insight, we establish simple scaling relations for allele frequency fluctuations, fixation probabilities, extinction times, and the site frequency spectra that arise when offspring numbers are distributed according to a power law $n^{-(1+\alpha)}$. To demonstrate that this coarse-grained model captures a wide variety of non-equilibrium dynamics, we validate our results in traveling waves, where the phenomenon of 'gene surfing' can produce any exponent $1 < \alpha < 2$. We argue that the concept of a dynamic sampling bias is useful generally to develop both intuition and statistical tests for the unusual dynamics of populations with skewed offspring distributions, which can confound commonly used tests for selection or demographic history.

I. INTRODUCTION

Interpreting the genetic differences between and within populations we observe today requires a robust understanding of how allele frequencies change over time. Most theoretical and statistical advancements have been based on the Wright-Fisher model [1, 2], which has shaped the intuition of generations of population geneticists for how evolutionary dynamics works [3]. The Wright-Fisher model assumes that the genetic makeup of a generation results from resampling the gene pool of the previous generation, whereby biases are introduced to account for most relevant evolutionary forces, such as selection, migration, or variable population sizes. For large populations, the resulting dynamics can be approximated by a biased diffusion process, which simplifies the statistical modeling of the genetic diversity. More importantly, Wright-Fisher diffusion is the limiting allele frequency process of a wide variety of microscopic models, as long as they satisfy seemingly mild assumptions (see below). This flexibility has made Wright-Fisher diffusion the standard model of choice to infer the demographic history of a species, loci of selection or the strength of polygenic selection [4–9].

Despite its versatility, Wright-Fisher diffusion can be a poor approximation when the population dynamics is driven by rare but strong number fluctuations. It is increasingly recognized that number fluctuations can be inflated for very different reasons. First, the considered species may have a broad offspring distribution, which occurs for marine species and plants with a Type III survivorship curve [10, 11] as well as viruses and fungi (reviewed in [12]). Broad offspring distributions also arise in infectious disease, when relatively few super-spreaders are responsible for the majority of the disease transmissions [13]. In the recent SARS-CoV-2 pandemic, for example, a strongly skewed offspring distributions were consistently inferred from both contact tracing data and infection cluster size distributions [14, 15]. Understanding allele frequency trajectories in these systems is extremely challenging, as statistical inference based on the Wright-Fisher model is often misleading (see e.g. [16]).

A second mechanism for strong number fluctuations are so-called jackpot events, which can occur in any species no matter the actual offspring distribution. Jackpot events are population bottlenecks that arise when the earliest, the most fit or the most advanced individuals have an unusual large number of descendants. Temporal jackpot events ("earliest") were first discovered by Luria and Delbrück [17] and studied as a signal of spontaneous mutations in an expanding population. They observed that a phage resistant mutant clone can grow exceptionally large if the resistance mutation by chance occurs early in an expansion event. Despite being rare, these jackpot events are easily detectable in large populations because they strongly inflate the variance of the expected number of mutants and lead to power-law descendant distributions.

The very same descendant distribution arises in models of rampant adaptation and of background selection. In these models, mutations generate jackpot events when they arise within the few fittest individuals [18]. Jackpot events also arise in range expansions, where the most advanced individuals in the front of the population have

a good chance to leave many descendants over the next few generations. This phenomenon of gene surfing can produce a wide range of scale-free descendant distributions [19–22].

To account for skewed offspring distributions, a number of theoretical studies have been conducted in the context of the coalescent framework. Based on this backward-in-time, a striking feature of broad offspring distributions is the simultaneous merging of multiple lineages. One of the most widely studied models is the beta-coalescent [23], which is a subclass of the Λ-coalescent and corresponds to the population dynamics with a power-law offspring number distribution $\propto u^{-1+\alpha}$. The case $\alpha = 1$, called Bolthausen-Sznitman coalescent [24], has been shown to be the limiting coalescent in models of so-called "pulled" traveling waves, which describe the most basic scenarios of range expansions [25] and of rampant adaptation [18, 26–28]. Moreover, so-called "semi-pushed" traveling waves that contain some level of co-operativity, induced e.g. by an Allee effect, generate power-law offspring distributions with $1 < \alpha < 2$ [21], indicating that their coalescent is intermediate between the Bolthausen-Sznitman and Kingman coalescents.

The tractability of coalescent approaches make it particularly useful for inferring demographic histories and detecting outlier behaviors [29–31]. However, as it is notoriously difficult to integrate selection in coalescent frameworks, there is also a strong need for forward-in-time approaches that capture the competition between genetic drift and selection. While for $\alpha \geq 2$, the limiting allele frequency dynamics is given by the well-understood Wright-Fisher process, much less is known for the case $\alpha < 2$. This is unfortunate because, as mentioned above, any exponent $1 \leq \alpha \leq 2$ can arise dynamically.

Recently, the forward-dynamics of the special case $\alpha = 1$ was studied by one of the authors [32], finding that an emergent sampling bias generates strong deviations from Wright-Fisher dynamics. The sampling bias arises because, in each generation, an allele with high frequency can sample more often and, hence, deeper into the tail of the offspring distribution than an allele with small frequency. The major allele of a biallelic site, therefore, has with high probability a greater number of offspring per individual than the minority type. This sampling bias acts like a selective advantage of the major allele, but its average effect is compensated by rare frequency hikes of the minor allele so that the expected change in frequency only changes in the presence of genuine selection.

Here, we focus on the understudied case $1 < \alpha < 2$ intermediate between the known cases of $\alpha \geq 2$, corresponding to Wright-Fisher diffusion, and $\alpha = 1$ described by jumps and sampling bias but vanishing diffusion. Similarly to the $\alpha = 1$ borderline case, we find that a minor-allele-suppressing sampling bias arises but that it is fading over time as the offspring distributions are sampled more and more thoroughly. This time-dependent sampling bias determines the scaling of the fixation probability, extinction time, stationary distribution, and site frequency spectrum. The combination of jumps and bias generates a so-called Levy-flight which controls the variability of allele frequency trajectories, for instance between unlinked genes or between populations. The flexibility of our model should enable to fit wide range cases that deviate from Wright-Fisher diffusion.

II. SAMPLING ALLELE FREQUENCIES ACROSS GENERATIONS

To study the impact of broad offspring numbers, we consider an idealized, panmictic, haploid population of constant size N that produces non-overlapping generations in the following way. First, we associate with each individual i a "reproductive value" [1, 33] U_i, which represents its *expected* contribution to the population of the next generation. The random numbers U_i are drawn from a specified distribution P_U. In a second step, we sample each individual in proportion to its reproductive value until we have obtained N new individuals representing the next generation.

Our model belongs to the general class of Cannings models [34]. The Wright-Fisher model is obtained if we choose P_U to be a Dirac delta function, such that all individuals have the same reproductive value.

We focus most of our analysis on the dynamics of two mutually exclusive alleles, a wild type and a mutant allele. The dynamics of the two alleles is captured by the time-dependent frequency $X(t)$ $(0 \leq X \leq 1)$ of mutants. The wild type frequency is given by $1 - X(t)$. The total reproductive values M and W of the mutant population and the the wild type population, respectively, are given by

$$M \equiv \sum_{i=1}^{NX} U_i^{(\mathrm{M})}, \quad W \equiv \sum_{i=1}^{N(1-X)} U_i^{(\mathrm{W})} . \tag{1}$$

Here, $U_i^{(\mathrm{M})}$ and $U_i^{(\mathrm{W})}$ are the individual reproductive values of mutants and wild types and sampled from the distribution P_U. The population at the next generation is generated by binomially sampling N individuals with success probability $\frac{M}{M+W}$. Mutations and selection are included as in the Wright-Fisher model. If the fitness of the mutant relative to the wild-type is $1 + s$, where s is the selection coefficient, and the forward- and back- mutation rates are μ_1 and μ_2 respectively, then the success probability is given by $\frac{(1-\mu_2)(1+s)M+(1-\mu_1)W}{(1+s)M+W}$.

For the offspring distribution P_U, we consider a family of fat-tailed distributions, which asymptotically behave as $P_U \sim \frac{1}{u^{\alpha+1}}$ with α being a positive constant. To make our presentation concrete, we choose $P_U(u) = \alpha/u^{\alpha+1}$ $(u \geq 1)$, which is known as the Pareto distribution. In the large population size limit, the neutral allele-frequency dynamics is known to only depend on the asymptotic power law exponent α provided we measure time in units of the coalescence time [35].

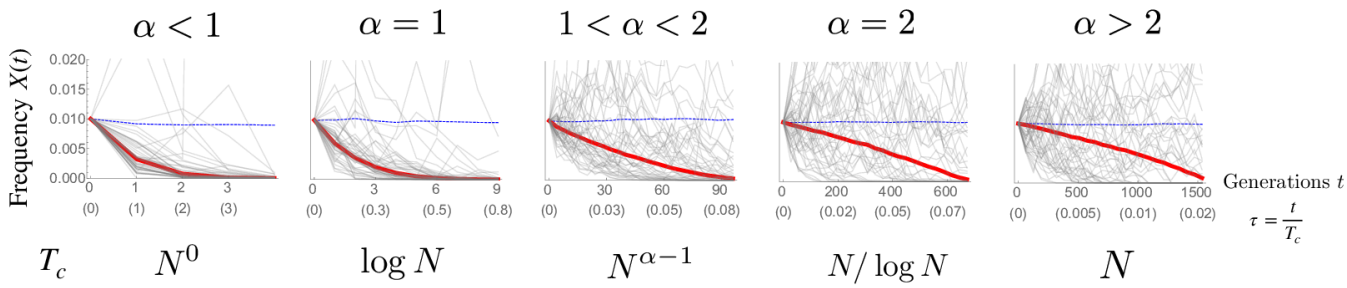

Figure 1. The mean (blue dashed curve) and median (red solid curve) of allele frequency trajectories for $\alpha = 0.8, 1, 1.5, 2$, and 2.5. For each α, 10^4 trajectories are generated with the initial frequency $x_0 = 0.01$. For ease of viewing, only 50 trajectories are shown in gray in each panel. The time t in units of generations and the one $\tau = \frac{t}{T_c}$ re-scaled by the coalescent timescale T_c are shown in the horizontal axes. The dependence of the coalescent time on the population size N is written below each panel. The population size is $N = 10^5$.

III. SIMULATION RESULTS

Our goal is to understand the asymptotic dynamics of our model for large N, where the frequency becomes continuous over time [36, 37] provided that $\alpha \geq 1$ [35]. We first present simulation results regarding relevant measures in the population genetics. Later, we provide a heuristic argument to explain them. Many separate observations (the fixation probability, extinction time, allele frequency fluctuations, stationary distribution, and site frequency spectrum) can be matched up with a unifying scaling picture.

Below, t and $\tau = t/T_c$ denote a time in units of generations and one normalized by the characteristic (coalescent) timescale T_c, respectively. T_c depends on the population size and the exponent α as follows: $T_c = N$ when $\alpha > 2$, $T_c = N/\log N$ when $\alpha = 2$, $T_c = N^{\alpha-1}$ when $1 < \alpha < 2$, and by $T_c = \log N$ when $\alpha = 1$. These timescales were originally derived in the coalescent framework [35]. Later, we explain how they can be rationalized within the forward-in-time approach.

To understand the frequency dynamics when $1 \leq \alpha < 2$, it is essential to distinguish between average and typical trajectories. As a proxy for typical trajectories, we use the median of the frequencies, denoted by $X^{\text{med}}(\tau)$, throughout this paper.

A. Neutral dynamics: typical trajectories and extinction time

First, we characterize the allele frequency dynamics in the absence of selection $s = 0$. In this neutral limit, the expected value of the allele frequency does not change over time, i.e., $\langle X(t) \rangle = X(0)$. Yet, despite the overall neutrality, a typical trajectory experiences a bias against the minority allele. This can be seen in Figure 1, where the mean and median are plotted across many realizations that start from the same frequency $X(0) = 0.01$. While the mean does not change over time, as required from neutrality, the median decays to zero in an α–dependent manner. By symmetry, the median increases towards fixation if the starting frequency is larger than 50%. Thus, the median experiences a bias against the minor allele. Note also that, when $1 < \alpha < 2$, the velocity of the median approaching extinction decreases as it approaches the extinction boundary (see the red curve in Figure 1). As we will show later, an uptick of the site frequency spectrum at the boundaries originates from this slowing.

Numerical simulations of the early part of trajectories show that time-dependent median displacement follows a simple power law,

$$\Delta X^{\text{med}} \equiv X^{\text{med}}(\tau) - X(0) \sim -\tau^{\frac{1}{\alpha}}, \tag{2}$$

up to a frequency-independent prefactor. Figure 2 shows the numerical result for $\alpha = 1.5$. The red curve represents the median of trajectories, which agrees well with $\Delta X^{\text{med}} \sim -\tau^{\frac{2}{3}}$.

Next we quantify the time to extinction, which turns out to be driven by the above minor-allele suppressing bias. Numerical results of the mean extinction time are consistent with

$$\tau_{\text{ext}} \sim X(0)^{\alpha-1}, \tag{3}$$

as shown in Figure 3. Hence, in units of the coalescence time, the mean extinction time τ_{ext} becomes larger as α decreases (namely, for a broader offspring distribution). Note, however, that if one measures time in units of generations, Equation 3 can be rewritten as $t_{\text{ext}} = \tau_{\text{ext}} T_c \sim (NX(0))^{\alpha-1}$, which becomes smaller as α decreases since $NX(0) \geq 1$.

B. Allele frequency fluctuations as a signature of broad offspring distributions

Next, we explore to what extent the spectrum of allele frequency fluctuations can provide a clue for identifying the exponent α of the offspring distribution. A deviation

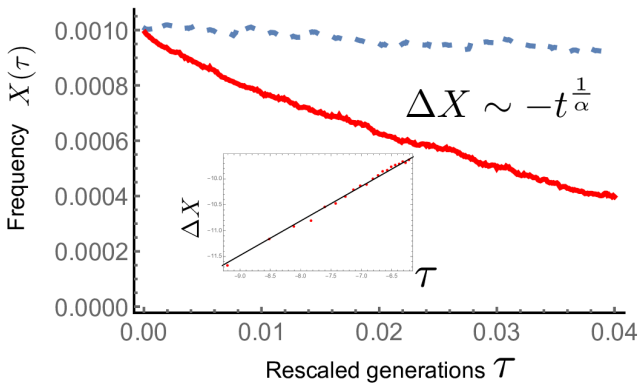

Figure 2. Trajectories of the median (red thick line) and the mean (blue dashed) of allele frequency when $\alpha = 1.5$ and $N = 10^8$. Inset: The trajectory (red points) of $|\Delta X^{\rm med}| = X(0) - X^{\rm med}(\tau)$ is shown in log-log plot. The median agrees well with the expectation from the scaling argument $\Delta X^{\rm med} \sim -\tau^{\frac{1}{\alpha}}$ (black solid line).

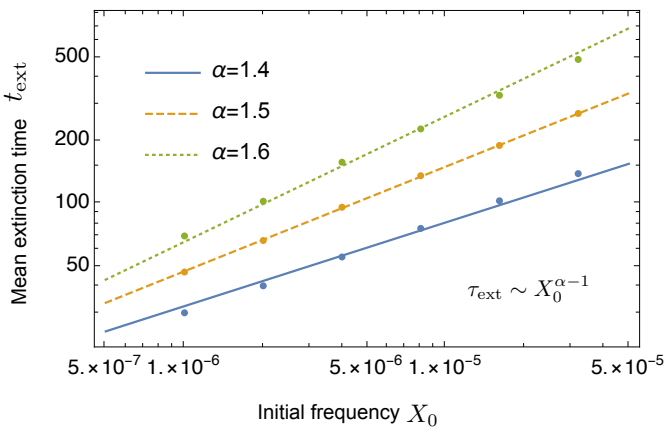

Figure 3. The mean extinction time $t_{\rm ext}$ (in units of generations) as a function of initial allele frequency $X(0)$ is plotted for $\alpha = 1.4$, 1.5, 1.6. Each of the straight lines has the slope $\alpha - 1$. $t_{\rm ext}$ can be fitted well by Equation 3. The population size is $N = 10^8$.

from the Wright-Fisher diffusion is most clearly revealed by measuring the median square displacement (median SD),

$$\text{Median SD} \equiv \mathbb{M}\left[\Big(X(t) - \mathbb{M}[X(t)]\Big)^2\right], \qquad (4)$$

where $\mathbb{M}[\cdot]$ denotes taking the median (e.g. $\mathbb{M}[X(t)] = X^{\rm med}(t)$). To measure the median SD, we simulate 1000 neutral allele frequency trajectories with initial condition $X(0) = 0.5$, for $\alpha = 1,\ 1.5$ and the Wright-Fisher model (Figure 4A). As shown in Figure 4B, the median SD computed from this data set is consistent with the scaling,

$$\text{Median SD} \sim t^{\frac{2}{\alpha}}, \qquad (5)$$

when $t/T_c \ll 1$. Noting $1 \le \alpha < 2$, this scaling means that typical fluctuations characterized by the median SD exhibit super-diffusion.

Usually, allele frequency fluctuations are quantified by using the mean SD $\equiv \langle (X(t) - X(0))^2 \rangle$, rather than the median SD. For the Wright-Fisher diffusion, the distinction between these two measures is irrelevant since both of them increase linearly with time, except with differing prefactors. However, for $1 \le \alpha < 2$, the α-dependence in Equation 5 can be detected by measuring the median SD. As shown in Figure 4C, the mean SD (computed from a large data set) grows linearly in time even when α is less than 2, as if the underlying process was diffusive.

That the dynamics is not diffusive also impacts the mean SD, but somewhat subtly in that its value depends on the size of the data set (i.e., the number of frequency trajectories) used to measure it. This is because while rare large jumps contribute the mean SD in a large data set, these jumps are not observed in a small data set (with high probability). To demonstrate this data-size dependence, we prepare an ensemble of data sets, where each data set consists of a given number of allele-frequency trajectories. Then, for each data set, we measure the diffusion exponent κ, which is defined by

$$\text{Mean SD} \propto t^{\kappa}. \qquad (6)$$

In Figure 4D, the ensemble-averaged exponent is shown by the blue circle. We can see that, as the data size increases, fluctuations characterized by the mean SD exhibit a crossover from super-diffusion ($\kappa = \frac{2}{\alpha}$) to normal diffusion ($\kappa = 1$). For the median SD, by contrast, we find that its diffusion exponent κ can be computed reliably without any significant dependence on the size of the data set (orange triangles in Figure 4D). For example, under the parameter setting in Figure 4D, given a date set of 320 trajectories, the diffusion exponent $\kappa_{\rm estimated}$ of the median SD falls within the interval $[1.45, 1.57]$ with probability $\sim 68\%$. This in turn predicts $\alpha_{\rm estimated} := \frac{\kappa_{\rm estimated}}{2} \approx 1.32 \pm 0.05$, which is close to the actual value $\alpha = 1.3$.

C. Fixation probability

Next, we examine the effect of natural selection on the fixation probability of beneficial mutations. We consider a mutant with positive selective advantage $s > 0$ arising in a monoclonal population. The fixation probability $P_{\rm fix}(s)$ of a single mutant depends on the parameter α of the offspring distribution. In the Wright-Fisher model (or equivalently, $\alpha \ge 2$), the fixation probability can be obtained using a diffusion approximation and is given by $P_{\rm fix}(s) = \frac{1-e^{-2s}}{1-e^{-2Ns}}$, which becomes $P_{\rm fix} \approx 2s$ when $Ns \gg 1$ and s is small. When $\alpha = 1$, an analytic result has been recently obtained in [32], which can be approximated as $P_{\rm fix}(s) \sim \frac{1}{N^{1-s}}$. For the intermediate case, $1 < \alpha < 2$, we find that the fixation probability is given

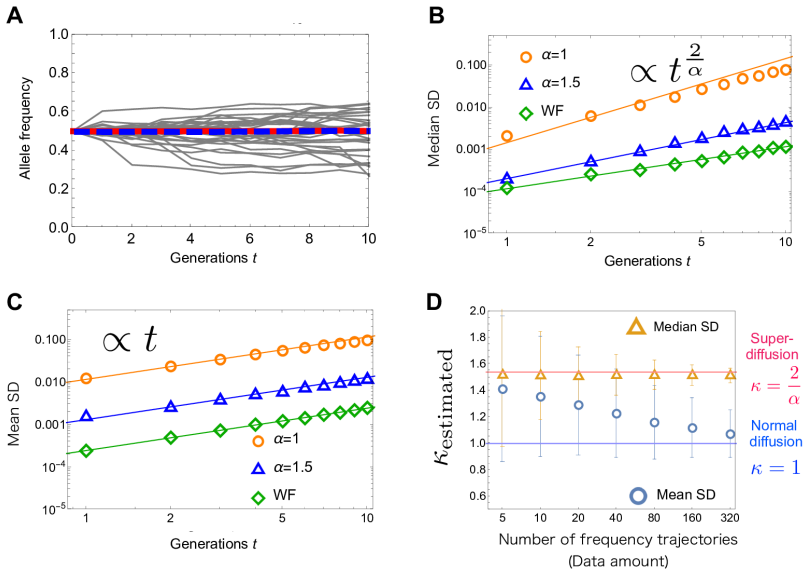

Figure 4. (**A**) Fluctuations of neutral allele frequencies when $\alpha = 1.5$ and $N = 10^5$. For $X(0) = 0.5$, the median $X^{\text{med}}(t)$ (red) is constant as well as the mean $\langle X(t) \rangle$ (blue). (**B**) The median square displacement computed from a data set of 1000 trajectories. For $\alpha = 1, 1.5$ and the Wright-Fisher model, $N = 10^8, 10^4$, and 10^3 are used respectively. The straight lines represent the scaling in Equation 5. For $\alpha = 1$, the fitting after $t \gtrsim 5$ is not perfect, since $t/T_c = t/\ln N \ll 1$ is not satisfied. (**C**) The mean square displacement (mean SD) for different values of α. Solid lines represent linear scaling, which is expected for a regular diffusion process. (**D**) Data-size dependence of the estimated diffusion exponent $\kappa_{\text{estimated}}$ for the mean SD (blue circle) and that for the median SD (orange triangle). See the main text for the detailed explanation. The horizontal lines show $\kappa = \frac{2}{\alpha}$ and $\kappa = 1$. The bars represent the standard deviations of $\kappa_{\text{estimated}}$. $\alpha = 1.3$ and $N = 10^8$ are used.

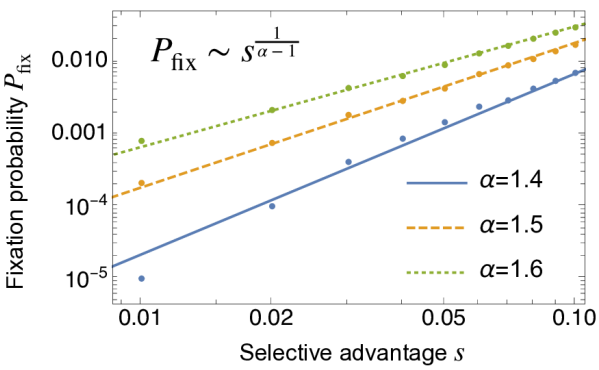

Figure 5. The fixation probability P_{fix} as a function of selective advantage s. The lines are the expectations from the scaling argument in Equation 42. The population size is $N = 10^8$.

by

$$P_{\text{fix}}(s) \sim s^{\frac{1}{\alpha-1}}. \tag{7}$$

See Figure 5 for the numerical results. Note that since $P_{\text{fix}}(s) \to \frac{1}{N}$ in the neutral limit independently of α, these results hold for sufficiently strong selection, $1 \ll N s^{\frac{1}{\alpha-1}}$.

As Equation 7 shows, for a fixed population size and selective advantage, the fixation probability becomes smaller as α decreases. Intuitively, this is because, for smaller α, the success of fixation in catching a ride on a jackpot event depends more strongly on luck than on fitness differences.

D. Site frequency spectrum

The site frequency spectrum (SFS) is often used as a convenient summary of the genetic diversity within a population. Theoretically, the SFS is defined in the infinite alleles model [38] as the density $f_{\text{SFS}}(x)$ of neutral derived alleles in the population (namely, $f_{\text{SFS}}(x)dx$ is the number of derived alleles in the frequency window $[x - \frac{dx}{2}, x + \frac{dx}{2}]$).

Figure 6 shows numerical plots of the SFS for $\alpha = 1, 1.5$, and the Wright-Fisher model. In the standard Wright-Fisher model, the SFS is proportional to $1/x$, which decreases monotonically as x increases. By contrast, when offspring numbers are broadly distributed (when $\alpha < 2$), the SFS is non-monotonic with a somewhat surprising uptick towards the fixation boundary. When $\alpha = 1$, the analytic understandings of asymptotic behaviors near both boundaries are well-established: $f_{\text{SFS}}(x)$ is proportional to $\frac{1}{(x \log x)^2}$ near $x \sim 0$ and $-\frac{1}{(1-x)\log(1-x)}$ near $x \sim 1$, respectively [18, 27] (see also Appendix E).

For the intermediate case $1 < \alpha < 2$, the rare-end behavior of the SFS has been analytically studied. From a backward approach (the Λ-coalescent), the authors in [39] showed

$$\lim_{n\to\infty} \frac{\zeta_i^{(n)}}{n^{2-\alpha}} \propto \frac{\Gamma(i+\alpha-2)}{i!}. \tag{8}$$

Here, n is a sample size and $\zeta_i^{(n)}$ is the number of sites at which variants appear i times in the sample (see [39] for the proportionality constant of the right-hand side of Equation 8). By using Stirling's approximation in Equa-

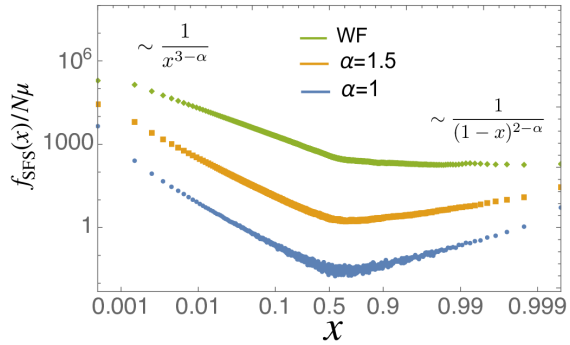

Figure 6. The site frequency spectrum for different values of α and fixed population size $N = 10^5$. When $1 < \alpha < 2$, the rare-end spectrum and the frequent-end spectrum are $\propto \frac{1}{x^{3-\alpha}}$ and $\propto \frac{1}{(1-x)^{2-\alpha}}$, respectively (see also Figure 7).

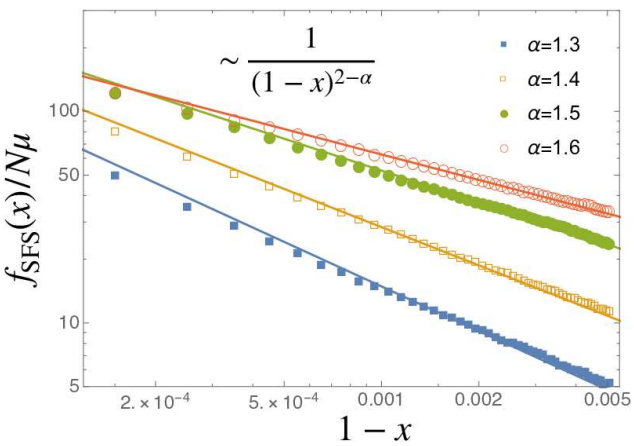

Figure 7. The SFS near $x = 1$ for $\alpha = 1.3,\ 1.4,\ 1.5,\ 1.6$ (circle, squares). The horizontal axis is $1 - x$. The solid lines are drawn assuming $f_{\rm SFS}(x) \propto \frac{1}{(1-x)^{2-\alpha}}$. $N = 10^6$ is used.

tion 8, we have

$$f_{\rm SFS}(x) \propto \frac{1}{x^{3-\alpha}} \quad \text{when } x \ll 1. \tag{9}$$

Equation 8, cannot be used for high-frequency variants, because the number of times the variants appear (i in Equation 8) is kept finite in taking the limit of the sample size n. To the best of our knowledge, a precise behavior at the high-frequency end for $1 < \alpha < 2$ has not been reported. As shown in Figure 7, we find that the asymptotic form of the uptick of $f_{\rm SFS}(x)$ is given by

$$f_{\rm SFS}(x) \propto \frac{1}{(1-x)^{2-\alpha}} \quad (\text{for } 1 - x \ll 1). \tag{10}$$

E. Mutation-drift balance

A broad offspring distribution also affects the stationary distribution of allele frequency when mutations and genetic drift are balancing one another. For simplicity, we consider symmetric reversible mutations between two neutral allele types. We denote the scaled mutation rate (per unit time in the continuous description) as $\theta = T_c\mu$, where μ denotes the mutation rate per generation. In the Wright-Fisher model, it is known that the stationary distribution is given by [36]

$$P_{\rm WF}(x) \propto x^{2\theta-1}(1-x)^{2\theta-1}. \tag{11}$$

There is a critical value $\theta_c^{\rm WF} = \frac{1}{2}$: When $\theta > \theta_c^{\rm WF}$, the distribution in Equation 11 has a single peak at the center $x = \frac{1}{2}$; when $\theta < \theta_c^{\rm WF}$, it has a U-shaped distribution, where the density is increasing monotonically from the center to the boundaries.

Figure 8A and B show the numerical results of the stationary distributions for the Wright-Fisher model and $\alpha = 1.5$, respectively. When $1 \le \alpha < 2$, while a critical value of the mutation rate θ_c exists as in the Wright-Fisher model, there is a qualitatively different feature: For a small mutation rate $\theta < \theta_c$, the stationary distribution is not a U-shaped but an M-shaped distribution with two peaks near the boundaries. Note that the M-shaped distribution indicates a stochastic switching behavior, as illustrated in Figure 8D) (the blue curve). As shown in Figure 8D, the peak positions are approximately given up to prefactors by

$$x_{\rm peak},\ 1 - x_{\rm peak} \approx \theta^{\frac{1}{2-\alpha}}. \tag{12}$$

In Appendix H, we show that the M-shaped stationary distribution persists even in the presence of natural selection, provided that selection is weaker than the sampling bias at the peaks of the distribution.

A similar M-shaped distribution was observed for the EW process in [40], wherein moments of the stationary distribution were extensively studied. However, the origin of the M-shaped distribution remained unclear. Below, using scaling arguments, we explain why the bimodal distribution arises in our case.

IV. ANALYTICAL ARGUMENTS

A. Limiting process, transition density, and time-dependent effective bias

We now provide analytical arguments for the observations made in the simulations described in the first part of this paper. Our discussion starts with an exact but somewhat unwieldy description of the allele frequency dynamics. We then show how exact short-time and intermediate time asymptotics can be derived and used to rationalize the sampling bias and the scaling laws discovered above.

The allele frequency dynamics can be fully characterized by the transition probability density $w_N(y|x)$ that the mutant frequency changes from x to y in one generation. Since one generation consists of random offspring

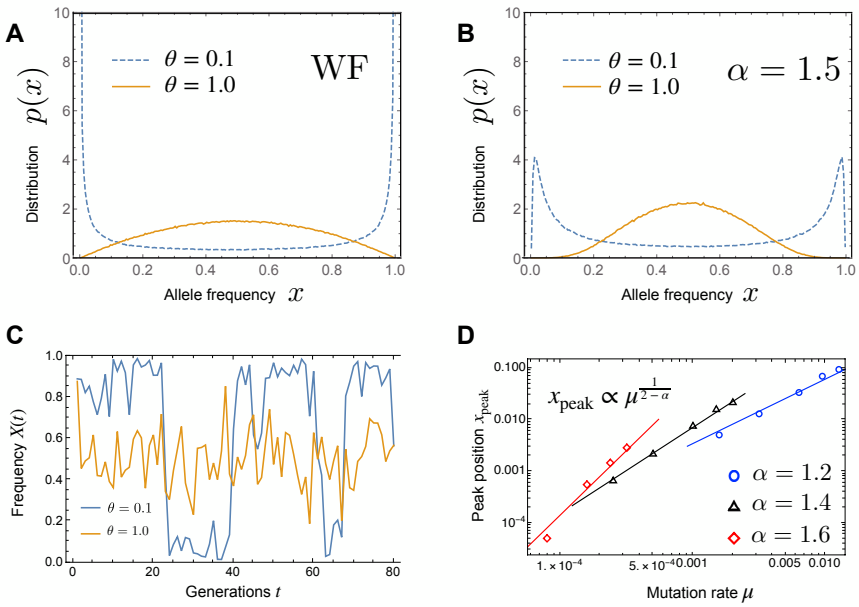

Figure 8. (**A**) Stationary distribution of the allele frequency in the Wright-Fisher model, when the mutation rate is small ($\theta = 0.1$) and large ($\theta = 1.0$). (**B**) Stationary distribution for $\alpha = 1.5$, when the mutation rate is small ($\theta = 0.1$) and large ($\theta = 1.0$). (**C**) The time-series of the allele frequency in the case of $\alpha = 1.5$, when the stationary distribution is bimodal ($\theta = 0.1$) and unimodal ($\theta = 1.0$). (**D**) The position of the peak near $x = 0$ of the stationary distribution versus the mutation rate μ. $N = 10^4$ is used.

contributions to the seed pool and binomial sampling from the seed pool, we have

$$w_N(y|x) = \int dM \int dW\, P^{\rm MUT}(M; xN) P^{\rm WT}(W; (1-x)N) \times {\rm Pr}_{\rm binom.}(yN, N, \frac{M}{M+W}). \tag{13}$$

Here, $P^{\rm MUT}(M; xN)$ is the probability density that the sum of xN random mutant offspring numbers takes the value M, $P^{\rm WT}(W; (1-x)N)$ is that for the wild type, and ${\rm Pr}_{\rm binom.}$ is the probability of getting yN successes in N trials with success probability $\frac{M}{M+W}$. First, we will focus on the neutral case, for which $P^{\rm MUT}$ and $P^{\rm WT}$ are the same function, i.e., $P^{\rm MUT}(\cdot) = P^{\rm WT}(\cdot)$.

While the resampling distribution w_N may in general behave in complex ways, it has few options in the large N limit. These constraints emerge from two asymptotic simplifications. First, since M and W are the sums of many random variables, $P^{\rm MUT}$ and $P^{\rm WT}$ tend to stable distributions as described by the generalized central limit theorem [41, 42] (see also Appendix A for a brief description of the theorem). Second, the fluctuations associated with binomial sampling become negligible compared with those induced by offspring number contributions to the seed pool, provided that the offspring distribution is sufficiently broad, i.e., $\alpha \leq 2$. Thus, we can replace ${\rm Pr}_{\rm binom.}(yN, N, \frac{M}{M+W})$ with a Dirac delta function, $\delta(y - \frac{M}{M+W})$. By using these facts and evaluating the integral in Equation 13 (see Appendix B for details), we obtain a simple analytical expression of $w_N(y|x)$, which is valid in the large N limit: When $\alpha = 1$ [32],

$$w_N(y|x) = \frac{1}{\log N} \frac{x(1-x)}{(x-y)^2}. \tag{14}$$

When $1 < \alpha < 2$,

$$w_N(y|x) = \begin{cases} N^{1-\alpha} C_\alpha\, x(1-x) \frac{(1-y)^{\alpha-1}}{(y-x)^{\alpha+1}} & \text{when } x < y \\ N^{1-\alpha} C_\alpha\, x(1-x) \frac{y^{\alpha-1}}{(x-y)^{\alpha+1}} & \text{when } x > y, \end{cases} \tag{15}$$

where $C_\alpha \equiv \alpha \left(\frac{\alpha-1}{\alpha}\right)^\alpha$.

To obtain the continuum description, we must appropriately scale the time t with the population size N [37]. The characteristic timescale (coalescent timescale) T_c can be read from the dependence of the transition density on N. [32] showed that, when $\alpha = 1$, the resulting limiting process is described by

$$\frac{\partial}{\partial \tau} P(x, \tau) = -\frac{\partial}{\partial x} V(x) P(x, \tau) + \int dx' (w(x|x') P(x', \tau) - w(x'|x) P(x, \tau)), \tag{16}$$

where the jump kernel $w(x'|x)$ is given by

$$w(x'|x) = \frac{x(1-x)}{(x-x')^2}, \tag{17}$$

and the advection (bias) term $V(x)$ is given by

$$V(x) = -{\rm P.V.} \int dx' (x'-x) w(x'|x) = x(1-x) \log \frac{x}{1-x}, \tag{18}$$

where P.V. denotes the Cauchy principal value. It is easy to check that Equation 18 satisfies the neutrality condition $\frac{\partial}{\partial \tau} \langle X \rangle = 0$.

To develop intuition, it is useful to interpret the different terms in Equation 16. First, $V(x)$ has a form of

frequency-dependent selection that enhances the major allele (with frequency $> 50\%$) and suppresses the minor allele. The apparent fitness differences between the mutant and wild type is given by the log-ratio of their frequencies. Such a selection-like effect arises because the major allele can sample the offspring number from $P_U(u)$ more deeply than the minor allele (see [32]). Second, in spite of this apparent bias, the neutrality of the whole process is maintained due to rare large jumps, characterized by $w(y|x)$. This also means that the neutrality does not hold if we focus on "typical" trajectories (see Figure 1). In fact, as we show in Appendix E, the median $x_{\rm med}$ of the mutant frequency, which is a proxy of "typical" trajectories, evolves according to

$$\frac{d}{d\tau}X^{\rm med}(\tau) = V(X^{\rm med}(\tau)) \quad (\text{when } \alpha = 1). \tag{19}$$

When $1 < \alpha < 2$, using the same reasoning as the derivation of Equation 16 and choosing $\tau = \frac{t}{C_\alpha N^{\alpha-1}}$, we can obtain the following differential Chapman-Kolmogorov equation,

$$\frac{\partial}{\partial\tau}P(x,\tau) = -\frac{\partial}{\partial x}V(x)P(x,\tau) + \int_{|x'-x|>\epsilon} dx'(w(x|x')P(x',\tau) - w(x'|x)P(x,\tau)) \tag{20}$$

where

$$w(x'|x) = \begin{cases} x(1-x)\frac{(1-x')^{\alpha-1}}{(x'-x)^{\alpha+1}} & \text{when } x < x' \\ x(1-x)\frac{x'^{\alpha-1}}{(x-x')^{\alpha+1}} & \text{when } x > x' \end{cases} \tag{21}$$

and

$$V(x) = -\int_{|x'-x|>\epsilon} dx'(x'-x)w(x'|x). \tag{22}$$

As in Equation 16, the advection term guarantees the neutrality of allele frequency. Equation 21 means that, when $x < \frac{1}{2}$, rightward jumps occur more frequently than leftward ones, and this tendency reverses when $x > \frac{1}{2}$. Noting the overall minus sign in Equation 22, this in turn means that $V_{\rm eff}$ is a bias against the minor allele (see Figure 1), as in the case of $\alpha = 1$. We will later show that when $x \ll 1$, the median trajectory is initially decaying like $\Delta X^{\rm med} \sim -\tau^{\frac{1}{\alpha}}$ (Equation 2).

Note that, in the limit $\epsilon \to 0$, two divergencies arise in Equation 20, one in the integral for the advection velocity in Equation 22 and one in the jump integral in Equation 16. However, since both divergencies exactly cancel, the entire right-hand side of Equation 20 is well-defined. As shown in Appendix D, Equations 16 and 20 can also be derived as a dual of the Λ-Fleming-Viot process, namely as the adjoint operator of the backward generator (e.g., [43, 44]).

Although it is difficult to study Equation 20 analytically, it is possible to derive exact short-time and long-time asymptotics that, combined with scaling arguments, paint a fairly comprehensive picture of the ensuing statistical genetics.

B. Short-time dynamics and fluctuations

First, we describe the transition density $P(x,\tau|x_0,\tau = 0)$ of Equation 20 for small times. When $1 < \alpha < 2$, the allele frequency changes due to the deterministic bias $V(x)$ and random occurrence of jumps, sampled from the broad distribution in Equation 21. Since the number of jump events is enormous ($\sim \frac{\tau}{\epsilon^\alpha}$) even for small τ, the generalized central limit theorem applies, and $X(\tau)$ is asymptotically distributed according to a stable distribution [41]. For a general stable distribution, its analytical expression is not available, and only its characteristic function can be expressed analytically. As we show in Appendix C, the random displacement $\Delta X(\tau) = X(\tau) - x_0$ can be expressed as

$$\Delta X(\tau) = X(\tau) - x_0 = \gamma(\tau, x_0)Z. \tag{23}$$

Here Z is sampled from the stable distribution $p(z)$ whose characteristic function $\langle e^{ikZ}\rangle \equiv \int dz e^{ikz}p(z)$ is given by

$$\langle e^{ikZ}\rangle = \exp\left[-|k|^\alpha\left(1 - i\beta(x_0)\tan\frac{\pi\alpha}{2}\text{sign}(k)\right)\right], \tag{24}$$

and the scale parameter $\gamma(\tau,x)$ and the skewness parameter $\beta(x)$ are respectively given by

$$\gamma(\tau,x) \equiv \tau^{\frac{1}{\alpha}}\left(\frac{\pi(x(1-x)^\alpha + (1-x)x^\alpha)}{2\Gamma(\alpha+1)\sin\frac{\pi\alpha}{2}}\right)^{\frac{1}{\alpha}}, \tag{25}$$

$$\beta(x) \equiv \frac{x(1-x)^\alpha - x^\alpha(1-x)}{x^\alpha(1-x) + x(1-x)^\alpha}. \tag{26}$$

Note that statistical properties of Z are independent of τ, and $\Delta X(\tau)$ depends on τ via the scale parameter $\gamma(\tau,x_0)$. As shown in Figure 9A, for small times, the transition density $P(x,\tau|x_0,\tau = 0)$ computed from the stable distribution agrees precisely with numerical simulation results in the discrete-time model. Our result can be regarded as a counterpart of the Gaussian approximation often employed for Wright-Fisher diffusion (see [9] and the references therein).

Now, we study the mean and median of the allele frequency using the short-time expression. The mean does not change in time since $\langle\Delta X(\tau)\rangle = \gamma(\tau,x_0)\langle Z\rangle = 0$, which is consistent with the neutrality. On the other hand, the median changes as

$$\Delta X^{\rm med}(\tau) = \gamma(\tau,x_0)Z^{\rm med}(x_0), \tag{27}$$

where $Z^{\rm med}(x_0)$ denotes the median of Z. $Z^{\rm med}(x_0)$ depends on x_0 via $\beta(x_0)$ (see Equation 24), and $Z^{\rm med}(x_0) \lesseqgtr 0$ for $x_0 \lesseqgtr \frac{1}{2}$. Equation 27 agrees with numerical simulations in the discrete-time model, while $X(\tau)$ is close to the initial frequency x_0 (see the red and black curves in Figure 9 (B)).

The scaling property $\Delta X(\tau) \propto \tau^{\frac{1}{\alpha}}$ in Equation 2 immediately follows from Equation 27, since $\gamma \propto \tau^{\frac{1}{\alpha}}$. This

scaling implies that there is a time-dependent bias driving the median of the allele frequency. Differentiating Equation 27 with respect to time gives

$$\frac{d}{d\tau}X^{\text{med}}(\tau) = V_{\text{eff}}(\tau) \tag{28}$$

where the *effective time-dependent bias* $V_{\text{eff}}(\tau)$ is given by

$$V_{\text{eff}}(\tau) \equiv \frac{\partial\gamma(\tau, x_0)}{\partial\tau} Z^{\text{med}}(x_0). \tag{29}$$

Near the boundaries $x = 0$ and $x = 1$, $V_{\text{eff}}(\tau)$ is approximately given by

$$V_{\text{eff}}(\tau) \approx \begin{cases} -k\frac{x_0^{\frac{1}{\alpha}}}{\tau^{1-\frac{1}{\alpha}}} & (x \ll 1) \\ +k\frac{(1-x_0)^{\frac{1}{\alpha}}}{\tau^{1-\frac{1}{\alpha}}} & (1-x \ll 1) \end{cases}, \tag{30}$$

where $k \equiv \frac{|Z^{\text{med}}(x_0=0)|}{\alpha}(\frac{\pi}{2\Gamma(\alpha+1)\sin\frac{\pi\alpha}{2}})^{\frac{1}{\alpha}}$ is a positive constant.

The advection term arises from a sampling bias. Intuitively, the time-dependent bias $V_{\text{eff}}(\tau)$ arises from a time-dependence of the largest sampled offspring number (Figure 10). To see this, consider a typical trajectory of the allele frequency starting from x. Up to a short time τ, only jumps from x to $y \in [y_-(\tau), y_+(\tau)]$ are likely to occur, where $y_-(\tau)$ and $y_+(\tau)$ can be estimated from

$$\tau \times \int_0^{y_-(\tau)} w(y|x)dy \sim 1, \quad \tau \times \int_{y_+(\tau)}^1 w(y|x)dy \sim 1. \tag{31}$$

These conditions give

$$y_-(\tau) \sim \frac{x}{1+(\frac{\tau(1-x)}{\alpha})^{1/\alpha}}, \quad y_+(\tau) \sim \frac{x+(\frac{\tau x}{\alpha})^{1/\alpha}}{1+(\frac{\tau x}{\alpha})^{1/\alpha}}. \tag{32}$$

Because these small jumps cancel a part of the bias $V(x)$ in Equation 22, the typical trajectory is then driven by the uncanceled residual part of the bias $V(x)$,

$$V'_{\text{eff}} \equiv -\int_{y\in[0,y_-(\tau)]\cup[y_+(\tau),1]} dy(y-x)w(y|x). \tag{33}$$

When $x \ll 1$, the dominant contribution to this integral is from $y \approx y_+(\tau)$. Using $w(y|x) \sim \frac{x}{(y-x)^{\alpha+1}}$ from the first line of Equation 21 and $y_+ - x \sim (\tau x)^{\frac{1}{\alpha}}$ from Equation 32, the above integral can be evaluated as $V'_{\text{eff}} \sim -\frac{x}{(y_+-x)^{\alpha-1}} \sim -\frac{x^{\frac{1}{\alpha}}}{\tau^{\frac{\alpha-1}{\alpha}}}$, which agrees with V_{eff} in Equation 30 for $x \ll 1$ (up to the factor κ). When $1-x \ll 1$, the dominant contribution to V'_{eff} is from $y \approx y_-(\tau)$ and can be evaluated in a similar way, reproducing V_{eff} in Equation 30 for $1-x \ll 1$.

Allele frequency fluctuations are inconsistent with Wright-Fisher diffusion. In the simulations, we found that, for $1 \le \alpha < 2$, allele frequency fluctuations are inconsistent with Wright-Fisher diffusion and characterized by super-diffusion with diffusion exponent $\frac{2}{\alpha}$ (see Equation 5). This finding is readily explained by the short-time asymptotic in Equation 23. Recalling statistical properties of Z are independent of τ, the median SD is given by

$$\text{Median SD} = \gamma(\tau, x_0)^2\mathbb{M}[(Z - Z^{\text{med}})^2] \propto \tau^{\frac{2}{\alpha}}. \tag{34}$$

This scaling can also be justified heuristically by noting that, for $1 < \alpha < 2$, the square displacement is dominated by large jumps. During time τ, an allele frequency $X(\tau)$ around x typically jumps to $y_\pm$ given in Equation 32. When $\tau \ll 1$, it is easy to see $|y_\pm - x| \sim \tau^{\frac{1}{\alpha}}$ with x-dependent prefactors. Because the median SD is dominated by the largest displacements, it can be evaluated as

$$\text{Median SD} \sim (y_\pm(\tau) - x)^2 \sim \tau^{\frac{2}{\alpha}}, \tag{35}$$

where $\tau = \frac{t}{T_c} \ll 1$ is assumed.

C. Long-time dynamics and extinction time

Above, we saw that at short times, allele frequencies carry out an unconstrained Levy flight. This random search process, however, gets distorted as soon as the allele frequency starts to get in reach of one of the absorbing boundaries. Interestingly, the dynamics then enters a universal intermediate asymptotic regime that controls both the characteristic extinction time as well as establishment times and fixation probabilities.

To see this, let us consider the extinction dynamics of a trajectory starting from a small frequency $x_0 \ll 1$ (Figure 4). At short times, we can apply the short-time asymptotics in Equations 28, 30. We expect Equations 28, 30 to break down when the displacement $\Delta X^{\text{med}}(\tau)$ computed from Equation 28 becomes comparable to x_0, which occurs at $\tau \sim \tau_{\text{short}} \equiv x_0^{\alpha-1}$. By taking a coarse-grained view, the rate of the frequency change in τ_{short} is roughly given by

$$\frac{\Delta X^{\text{med}}}{\tau_{\text{short}}} \sim -x_0^{2-\alpha}. \tag{36}$$

This suggests that, in a long timescale ($\tau \gtrsim \tau_{\text{short}}$), the median frequency decreases as

$$\frac{d}{d\tau}X^{\text{med}}(\tau) = \tilde{V}_{\text{eff}}(X^{\text{med}}(\tau)) \quad (\text{for } X \ll 1), \tag{37}$$

where, up to a prefactor, the frequency-dependent bias $\tilde{V}_{\text{eff}}(X)$ is given by

$$\tilde{V}_{\text{eff}}(X) \sim -X^{2-\alpha}. \tag{38}$$

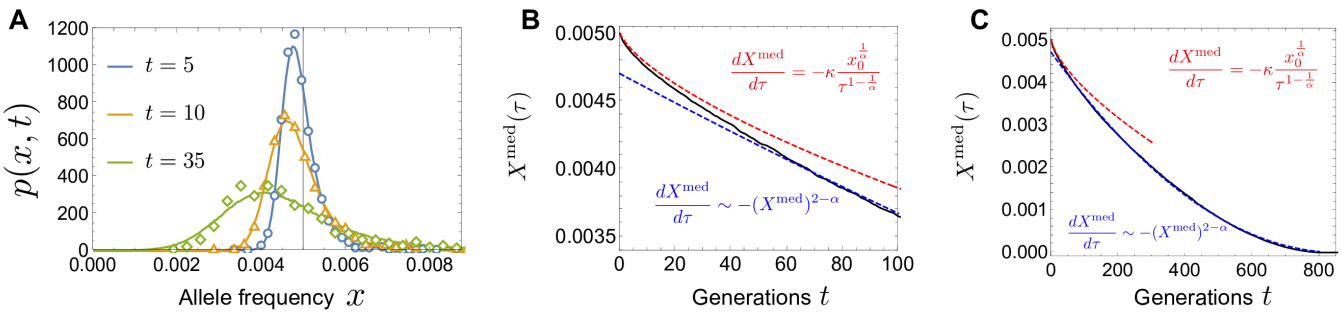

Figure 9. (A) The allele frequency distribution $p(x, t|x_0 = 0.005)$ at generation $t = 5, 10, 35$, for $\alpha = 1.5$. The solid lines denote the short-time transition densities given by Equations 23 and 24, and the open markers denote those computed from 10000 allele frequency trajectories in the discrete-time model. (B) The initial dynamics of the median of the allele frequency (black). The red and blue lines denote the short-time solution in Equation 27 and the long-time solution in Equation 37, where constants of integration and the prefactor of Equation 38 are determined by fitting to the discrete-time model (black line) between $40 < t < 800$. (C) The overall trajectory to extinction. The color scheme is the same as that in (B). In (A-C), $\alpha = 1.5$, $N = 10^7$, $x_0 = 0.005$ are used.

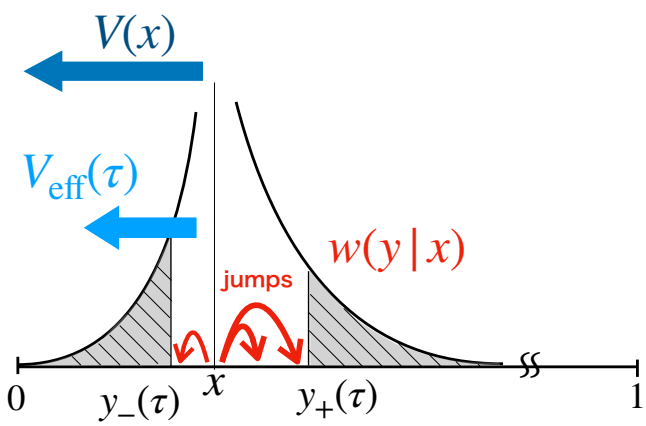

Figure 10. Schematic explanation of the effective time-dependent bias $V_{\rm eff}(\tau)$. The black curve shows the jump rate $w(y|x)$ in Equation 21 when $x \ll 1$. In a time τ, small jumps within the region $[y_-(\tau), y_+(\tau)]$ are likely to occur, offsetting a part of the original bias $V(x)$. $V_{\rm eff}(\tau)$ is the residual part of the bias.

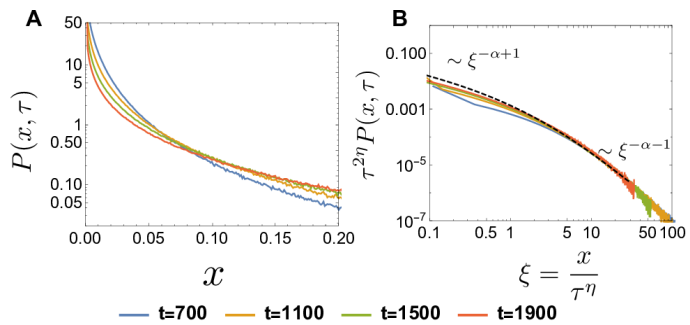

Figure 11. (A) Log plot of $P(x, \tau|x_0 = 0.01)$ at generations $t = 700, 1100, 1500, 1900$ computed from the discrete-time model. $N = 10^7$ and $\alpha = 1.5$. (B) Log-Log plot of $\tau^{2\eta}P(x, \tau|x_0 = 0.01)$ versus $\xi = x/\tau^{2\eta}$, where $\eta = (\alpha - 1)^{-1}$, at $t = 700, 1100, 1500, 1900$ (solid curves). The dashed curve represents the analytic result of $g(\xi)$ (see Appendix C). The curves $\tau^{2\eta}P(x, \tau|x_0 = 0.01)$ at the different time points collapse into $g(\xi)$, supporting the scaling ansatz in Equation 39.

In Figure 9C, it is numerically shown that the long-time trajectory $X^{\rm med}(\tau > \tau_{\rm short})$ is consistent with Equation 37. By solving Equation 37, the median trajectory goes to extinction at $\tau_{\rm ext} \sim x_0^{\alpha-1}$ (Equation 3), in agreement with our simulations (Figure 4). Note that, for $1 - x \ll 1$, the bias in Equation 38 is replaced by $\tilde{V}_{\rm eff}(X) \sim +(1 - X)^{2-\alpha}$.

Importantly, Equations 37 and 38 can also be rigorously justified from a scaling ansatz for the transition density. After some time, $P(x, \tau|x_0)$ spreads broadly over the region $x \ll 1$ with a peak at $x = 0$ (Figure 11A). As shown in Figure 11B, $P(x, \tau)$ is consistent with the following scaling ansatz;

$$P(x, \tau) \sim \tau^{-2\eta} g(\xi) \quad (\text{for } x \ll 1), \tag{39}$$

where $\eta \equiv \frac{1}{\alpha-1}$ and $g(\xi)$ is a function of $\xi \equiv \frac{x}{\tau^\eta}$. Up to an overall constant, $g(\xi)$ can be determined analytically and expressed as an infinite series (see Appendix C). Note that the τ-dependent factor in Equation 39 is motivated from the fact that the extent over which the distribution spreads increases like τ^η. Equation 39 implies that, conditional on establishment at τ, the median frequency increases as $X^{\rm med}(\tau)|_{\rm establish} \sim \tau^\eta$. Then, Equation 38 follows by evaluating the bias in Equation 30 at $\tau \sim (X^{\rm med})^{\frac{1}{\eta}}$ and at $X^{\rm med}$, instead of at x_0.

As a consistency check of the exponent $\alpha - 1$ in Equation 3, we consider two solvable, extreme cases. First, in the limit $\alpha \to 2$, the dependence on x_0 in Equation 3 becomes linear. In the Wright-Fisher model, the mean extinction time can be obtained analytically by solving the backward equation $-1 = x(1 - x)\frac{\partial^2 \tau_{\rm ext}(x)}{\partial x^2}$ (see, for example, [45]). The solution is proportional to x_0 with a logarithmic correction, $\tau_{\rm ext} \approx -x_0 \log x_0$. Second, when

$\alpha \to 1$, the mean extinction time no longer depends on x_0. We can obtain this explicitly, by solving Equation 19: Using $V(X) \simeq X \log X$ when $X \ll 1$, the solution is given by $\log \frac{\log X(\tau)}{\log x_0} = \tau$. Therefore, if we approximately define the mean extinction time $\tau_{\rm ext}$ as $X(\tau_{\rm ext}) \approx \frac{1}{N}$, we obtain $\tau_{\rm ext} \approx \log \frac{\log N}{-\log x_0} \approx \log\log N$, which is to leading order independent of x_0 if x_0 is taken to be of order one.

D. Natural selection and fixation probability

One important advantage of the forward-time perspective is that we account for natural selection by introducing an appropriate bias favoring of the beneficial variant. Suppose that the mutant type has a selective advantage $s > 0$, such that the average offspring number of mutants is increased by a factor of $1+s$ relative to the wild type. In time-rescaled Chapman-Kolmogorov equation, this adds the term $\sigma x(1-x)$, where $\sigma = T_c s$, into the advection $V(x)$ of Equation 20.

The key observation underlying the argument below is that when X is sufficiently small, the selection force is negligible compared to the bias $\tilde{V}_{\rm eff}(X)$ in Equation 38 because while the former is approximately linear in X, the latter is sublinear. If the frequency happens to grow and reach a certain value X_c, the genuine selection begins to dominate over the bias, and the trajectory fixes with high probability (see Figure 12 for example trajectories and Figure 13)A). By using Equation 38, the crossover point X_c can be estimated from balancing selection with the sampling bias,

$$\sigma\ X = -\tilde{V}_{\rm eff}(X) \sim X^{2-\alpha}, \tag{40}$$

which gives

$$X_c \sim \sigma^{-\frac{1}{\alpha-1}} = \frac{1}{N} s^{-\frac{1}{\alpha-1}}. \tag{41}$$

The fixation probability $P_{\rm fix}$ can be estimated by using the neutral fixation probability in a population of size $\approx NX_c$, because the dynamics are essentially neutral for $X \ll X_c$, and the trajectory grows almost deterministically for $X > X_c$. Thus, the fixation probability is approximately given by

$$P_{\rm fix} \sim \frac{1}{NX_c} = s^{\frac{1}{\alpha-1}}, \tag{42}$$

which is valid for $\frac{1}{N} \ll s^{\frac{1}{\alpha-1}}$. Equation 42 reproduces our simulation results in Figure 5 for $1 < \alpha < 2$ and, as $\alpha \to 2$, also reproduces the known result of the Wright-Fisher model, $P_{\rm fix}^{\rm WF} \sim 2s$ (up to a prefactor).

E. Site frequency spectrum

By using the time-dependent effective bias, we can also estimate the behavior of the SFS $f_{\rm SFS}(x)$ for frequent and rare variants. While the SFS is theoretically defined in the infinite alleles model, it can be computed from our biallelic framework ([46]): $f_{\rm SFS}(x)\Delta x$ is defined as the expected number of neutral derived alleles in the frequency interval $[x - \frac{\Delta x}{2}, x - \frac{\Delta x}{2}]$ in a sampled population (here, the whole population). Because new mutations are assumed to arise uniformly in time, the SFS for unlinked neutral loci is given by the product of the total mutation rate μN and the mean sojourn time, namely, the average time an allele spends in the frequency interval $[x - \frac{\Delta x}{2}, x - \frac{\Delta x}{2}]$ until fixation or extinction.

First, we consider the low-frequency end, $x \ll 1$, of the SFS (see [47] for a similar argument). Since the SFS is proportional to the sojourn time, trajectories whose maximum frequencies are x or slightly larger than x dominantly contribute to the SFS $f_{\rm SFS}(x)$ at x. Since these trajectories typically go extinct due to the bias, and we can roughly estimate their sojourn times at x as the inverse of "velocity", $|\tilde{V}_{\rm eff}(x)| \sim x^{2-\alpha}$ in Equation 38. Since the probability that a trajectory grows above a frequency x is roughly given by $\sim 1/(Nx)$, the SFS is

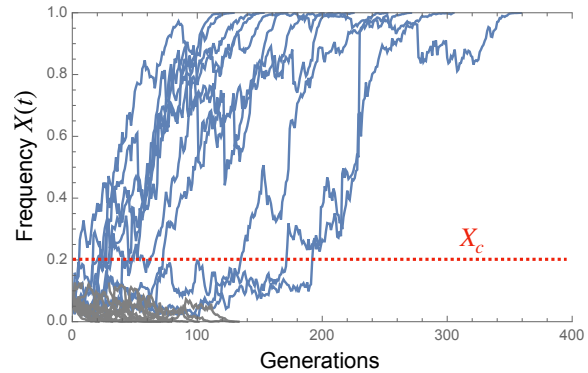

Figure 12. Example of trajectories of the frequency of the beneficial allele, starting from $x_0 = 0.05$. $\alpha = 1.5, s = 0.03$ and $N = 5000$. Fixed trajectories are colored in blue and extinct ones in gray. Here, the crossover point X_c can be estimated as $X_c \sim 0.2$ (, assuming that the proportional constant in Equation 41 is one). Once a trajectory reaches the crossover point, it becomes fixed in high probability.

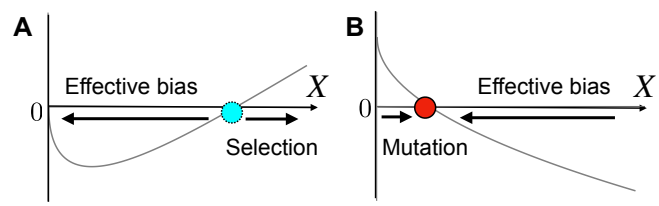

Figure 13. (A) The crossover from the effective bias to genuine selection. $V(X) \sim -CX^{2-\alpha} + \sigma X$ is plotted, where C is a positive coefficient and $\sigma > 0$. Deterministically, an unstable point exists at $x \sim X_c$. (B) The balance between the effective bias and mutation. $V(X) \sim -CX^{2-\alpha} + \theta$ is plotted. Deterministically, a stable point exists at $X \sim \theta^{1/(2-\alpha)}$.

proportional to

$$\frac{1}{Nx\tilde{V}_{\text{eff}}(x)} \propto \frac{1}{x^{3-\alpha}} \quad (\text{for } x \ll 1). \tag{43}$$

Similarly, for the high-frequency end of the SFS, only the trajectories that grow above x can contribute to $f_{\text{SFS}}(x)$. Typically, these trajectories go to fixation due to the bias $\tilde{V}_{\text{eff}}(x) \sim (1-x)^{2-\alpha}$. Therefore, the SFS is proportional to

$$\frac{1}{Nx\tilde{V}_{\text{eff}}(x)} \propto \frac{1}{(1-x)^{2-\alpha}} \quad (\text{for } 1-x \ll 1). \tag{44}$$

The effect of the genuine selection on the SFS can also be studied by using the effective bias. See Appendix F.

F. Bimodality of stationary distribution

Now, we turn to explaining the bimodality observed at mutation-drift balance. We found that, when the mutation rates are small, the stationary allele frequency distribution is not a U-shaped, as expected from Wright-Fisher dynamics, but M-shaped, as shown in Figure 8. The M-shaped distribution arises from the balance between the mutational force and the effective bias (see Figure 13B). In the Chapman-Kolmogorov equation, the mutational force is given by

$$-\theta x + \theta(1-x) \approx \begin{cases} +\theta & (x \ll 1) \\ -\theta & (1-x \ll 1), \end{cases} \tag{45}$$

which pushes the frequency toward the center $x = \frac{1}{2}$. On the other hand, the effective bias, $\tilde{V}_{\text{eff}}(x) \approx -x^{2-\alpha}$ for $x \ll 1$ and $\tilde{V}_{\text{eff}}(x) \approx (1-x)^{2-\alpha}$ for $1-x \ll 1$, pushes a trajectory toward the closer boundary. Therefore, the positions where these two forces balance are approximately given by

$$x_{\text{peak}} \approx c\theta^{\frac{1}{2-\alpha}}, 1 - c\theta^{\frac{1}{2-\alpha}}, \tag{46}$$

where c is a positive constant. If θ is sufficiently small, we can always find the balancing points. The presence of these two balancing points means that we can think of the allele frequency dynamics as a two-state system, essentially analogous to a super-diffusing particle in a double-well potential (see Figure 8C for a realization of trajectories). This explains the bimodal shape of the stationary distribution.

Finally, we remark that, even in the presence of natural selection, the balancing positions are still determined from the mutation-effective bias balance provided that $\theta \ll 1$: while the effective bias and the mutational term are sub-linear and constant respectively, the selection term $\sigma x(1-x)$ is linear in x when $x \ll 1$. Thus, when θ is sufficiently small, the magnitude of the selection term around $x = c\theta^{\frac{1}{2-\alpha}}, 1 - c\theta^{\frac{1}{2-\alpha}}$ is negligible, and the peak positions are given by Equation 46.

V. DISCUSSION

In this study, we analyzed the effect of power law offspring distributions on the competition of two mutually exclusive alleles. Our main reason to consider such broad offspring distributions is that they often emerge in evolutionary scenarios that inflate the reproductive value [33] of a small set of founders. For example, range expansions blow up the descendant numbers of the most advanced individuals in the front of the population, an effect that has been called gene surfing [19]. Likewise, continual rampant adaptation boosts the descendant numbers of the most fit individuals. The resulting allele frequency dynamics becomes asymptotically similar to that of a population with scale-free offspring distributions.

In the case of narrow offspring distributions, which is predominant assumption in population genetics, it is usually an excellent approximation to describe the allele frequency dynamics by a biased diffusion process, which forms the basis of powerful inference frameworks [9]. If the offspring distribution is broad, however, allele frequency trajectories are disrupted by discontinuous jumps, resulting from so-called jackpot events - exceptionally large family sizes drawn by chance from the offspring distribution. Our goal was to find an analytical and intuitive framework within which we can understand the main features of these unusual dynamics.

We found that the main counter-intuitive features can be understood and well-approximated from a competition of selection and mutations with a time-dependent emergent sampling bias, $V_{\text{eff}}(\tau)$. The sampling bias favors the major allele and arises, because the subpopulation carrying the major allele typically samples deeper into the tail of the offspring distribution than the minor allele fraction.

In the remainder, we first summarize the unusual population genetic patterns that can be explained by the action of these effective forces. We then discuss how broad offspring dynamics could be detected in natural populations and what its implications are for the dynamics of adaptation. Finally, we demonstrate that these dynamics are also ubiquitous in populations with narrow offspring distributions, when mutational jackpot are possible. Therefore, we believe our theoretical framework may be taken as a general null model for populations far from equilibrium.

A. Unusual dynamics

We found that the sampling bias effectively acts like time- and frequency-dependent selection. In the absence of true selection, $V_{\text{eff}}(x, \tau)$ drives the major allele to fixation, first rapidly and than gradually slowing down with time and proximity to fixation. The slowing down of the sampling bias near fixation also leads to an excess of high-frequency alleles, given continual influx of neutral mutations. This generates a high-frequency uptick

in the site frequency spectrum, which is characteristic of the tail of the offspring distribution. In mutation-drift balance, the allele frequency distribution is M-shaped, in contrast to the U-shape expected from Wright-Fisher dynamics. The peaks reflect the balance of the mutational and sampling bias.

Non-neutral dynamics depends on whether the genuine selection force dominates over the sampling bias. The sampling bias tends to dominate near extinction or fixation, and wanes near 50% frequency. A de-novo beneficial allele will not be able to fix unless it overcomes, by chance, the switch-point frequency at which genuine selection becomes stronger than the sampling bias. Finally, fluctuations in typical trajectories are getting stronger over time. As a consequence, allele frequencies super-diffuse: fluctuations grow with time more rapidly than under regular Wright-Fisher diffusion.

B. Detecting dynamics driven by broad offspring distributions

The time-dependent over-dispersion is most readily detected by plotting the median square displacement as a function of time (see Figure 4B). Testing deviations in this statistics are an attractive avenue for detecting deviations from Wright-Fisher diffusion because the signal is strong for intermediate allele frequencies, which can be accurately measured by population sequencing. By contrast, the time-dependent bias vanishes when an allele has 50% frequency. So, the detection of the sampling bias requires accurate time series data of low frequency variants, which is difficult to obtain given sequencing errors.

It is clear that a single super-diffusing but neutral allele would not abide by the diffusive Wright-Fisher null model and thus might be falsely considered as an allele under selection. But importantly, allele super-diffusion has an impact even on statistics that sum over many unlinked loci. This is significant for inference methods, for instance to detect polygenic selection, which argue that trait values follow a diffusion process, if not for an underlying Wright-Fisher dynamics of the allele frequencies then because they sum over many independent allele frequencies [6]. However, $\alpha < 2$ dynamics breaks both of these arguments. In particular, sums of many unlinked loci tend to non-Gaussian distributions (so called alpha-stable distributions). Hence, for traditional inference methods based on Wright-Fisher diffusion or standard central limit theorem [9], an underlying super-diffusion process should be ruled out.

If time series are not available, broad offspring numbers can also be detected from the site frequency spectrum (SFS) [18]. A tail-tale sign of the sampling bias is a characteristic uptick at the high-frquency tail of the SFS, which is difficult to generate by demographic variation [18]. As we have shown, the shape of the uptick is characteristic of the tail of the offspring distribution (the parameter α).

C. Implications for the dynamics of adaptation

We found that the fixation probabilities quite sensitively depends on the broadness α of the offspring distribution (Equation 42). Accordingly, the dynamics of adaptation, which ultimately depends on the fixation of beneficial variants, should change quantitatively. To estimate these modifications, we consider an asexual population of constant size N with a broad offspring distribution with $1 < \alpha < 2$, wherein beneficial mutations occur at the rate μ_{B}. For low mutation rates, mutations sweep one after the other but when mutation rate are sufficiently high, multiple mutations occur and most mutations are outcompeted by fitter mutations. Such a situation is known as clonal interference.

We can study the effect of the exponent α on the adaptation dynamics quantitatively by repeating the argument in [48], wherein the variance of offspring numbers is assumed to be narrow. As discussed in Appendix G, clonal interference should occur if

$$\mu_{\mathrm{B}} N s^{\frac{2-\alpha}{\alpha-1}} \ln(N s^{\frac{1}{\alpha-1}}) \gtrsim 1 \quad \text{(clonal interference)}, \tag{47}$$

where $s > 0$ is the fitness effect of a mutation, which we assume to be constant. The rate R of adaptation is given by

$$R \sim \begin{cases} \mu_{\mathrm{B}} N s^{\frac{\alpha}{\alpha-1}} & \text{(successive selective sweeps)} \\ \frac{2 s^2 \ln(N s^{\frac{1}{\alpha-1}})}{(\ln \frac{s}{\mu_{\mathrm{B}}})^2} & \text{(clonal interference)} \end{cases} \tag{48}$$

Note that the second line in Equation 48 reproduces Equation 5 of [48] in the limit $\alpha \to 2$. Thus, the rate of adaptation depends only weakly (logarithmically) on α in the clonal interference regime, even though the condition for clonal interference in Equation 47 depends on α quite sensitively, .

D. Emergence of skewed offspring distributions in models of range expansions

Our study can be regarded as an analysis of the population genetics induced by power-law offspring distributions. The main reason to consider these scale-free offspring distributions is that they quite generally *emerge* in models of stochastic traveling waves [21]. Such models are ubiquitous in population genetics because they describe a wide range of evolutionary scenarios, including range expansions, rampant asexual and sexual adaptation as well as Muller's ratchet [18, 21, 25–28]. Our analysis should therefore apply most directly to these evolutionary scenarios, which we now demonstrate using a simple model of a range expansion. We end by discussing the question of whether some of our results may also arise in scale-rich offspring distributions.

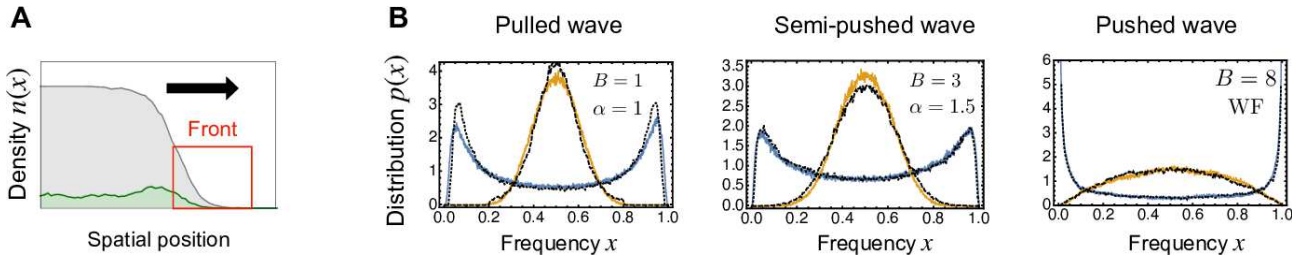

Figure 14. (**A**) The model of a range expanding population with two neutral alleles (green and gray). A broad offspring distribution arises dynamically in the front region. (**B**) Stationary distributions of the allele frequency when mutation rate θ is small (blue) and when θ is large (orange). The wiggling lines (blue/orange) are the numerical results in the traveling wave model, while the dotted lines (black) are those in the macroscopic model. The parameters of the Allee effect B are $B = 1$ (left), 3 (middle), and 8 (right). See Appendix I for the details of the implementation of the simulation and other parameter values.

Ref. [21] argued that any exponent $1 \leq \alpha \leq 2$ can emerge in a simple model of range expansions that incorporates a tunable level of cooperativity between individuals (Figure 14A). The model can be described by a generalized stochastic Fisher-Kolmogorov equation

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} + r(n)n + \text{noise}, \tag{49}$$

for the time-dependent population density $n(x,t)$ at position x in a linear habitat and time t. The growth rate $r(n)$ is assumed to be density-dependent, with

$$r(n) = r_0(1 - \frac{n}{K})(1 + B\frac{n}{K}), \tag{50}$$

where the parameter $B \geq 0$ accounts for co-operativity among individuals, which is also called an Allee effect. As discussed in [32], lineages in the region of the wave tip are diffusively mixed within the timescale $\tau_{mix} \sim \frac{1}{r}\ln^2 K\sqrt{\frac{D}{r}}$. This implies that, in this microscopic model, resampling from an offspring distribution roughly occurs every τ_{mix} generations. In [21, 22], it was argued that depending on the strength of the Allee effect, the offspring distributions corresponding to any of the three distinct classes of the beta coalescent process can arise; namely, the Bolthausen-Sznitman coalescent when $B < 2$, the beta coalescent with $1 < \alpha < 2$ when $2 < B < 4$, and the Kingman coalescent when $B > 4$.

To demonstrate clearly that our present study can serve as a macroscopic analysis of the traveling model, we introduce reversible mutations in the traveling wave model and measured the mutant frequency of the first $N \sim \frac{K}{k}$ individuals from the edge of the front. Here, k is the spatial decay rate, i.e., $n \sim e^{-k\tilde{x}}$ where $\tilde{x}$ is the coordinate comoving with the expansion. This definition of the mutant frequency is reasonable because only the wave front has a skewed offspring distribution due to the founder effect. In Figure 14B, for $B = 1$ (left), 3 (middle), and 8 (right), the frequency distributions in the traveling wave model are shown when the mutation rate is small (orange jagged line) and when it is large (blue jagged line). The corresponding distributions in the macroscopic model are shown by black dotted lines. The stationary distributions in the traveling wave model agree well with those in the macroscopic model. Especially, the transition from the M-shaped or U-shaped distribution to the monomodal distribution is consistently reproduced in the traveling wave model. These results underscore the correspondence between the traveling wave with the Allee effect and the beta coalescent process.

The above-described correspondence suggests that the spatial area occupied by one allele type in a range expansion should behave statistically like the time-integral over the allele frequency in the Cannings model. In the context of adapting (non-spatial) populations, this quantity describes the total number of mutational opportunities of a mutant lineage [48–50]. As presented in Appendix J, the distribution of the time-integrated frequency exhibits a scaling behavior, that depends on the offspring distribution sensitively. While a full discussion is beyond the scope of this paper, we expect that the distribution of areas serves as a useful observable to distinguish different prototypes of traveling waves [21].

Broad offspring distributions with a scale: While scale-free offspring distributions often emerge over an intermediate time scale (τ_{mix} in the above traveling wave model), there are also species that over single generations show broad offspring numbers and violate Wright-Fisher diffusion. For such species, it may be more natural to consider offspring distribution with a characteristic scale. In 'sweepstake' reproduction [11], a fixed and finite fraction of the population is replaced at every sweepstake event (specified by the parameter Ψ in [11]). Because Ψ sets a characteristic scale in offspring numbers, power law relationships for the median of allele frequencies as well as frequency fluctuations cannot be expected, which we confirm in Appendix K. Nevertheless, the qualitative features of a sampling bias can be recognized quite clearly for sweepstake reproduction as well.

Either type of model ultimately is an approximation to true offspring distributions, and it depends on the situation, which one to use. As we argued, the beta-coalescent along with the forward-in-time model described in this paper is the natural choice for range expansions, rapid adaptive process or other scenarios where the reproduc-

tive value of a chosen few are highly inflated.

ACKNOWLEDGEMENTS

This work is in part supported by RIKEN iTHEMS Program. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award R01GM115851, a National Science Foundation CAREER Award (#1555330), a Simons Investigator award from the Simons Foundation (#327934), and JSPS KAKENHI (Grant Number JP19K03663). We express our sincere thanks to Benjamin H. Good, Daniel B. Weissman, Jiseon Min, Joao Ascensao, Michael M. Desai, and Stephen Martis for their helpful discussions and comments.

[1] RA Fisher. The genetical theory of natural selection oxford university press. *London, UK*, 1930.

[2] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.

[3] James F Crow and Motoo Kimura. An introduction to population genetics theory. *An introduction to population genetics theory*, 1970.

[4] Jonathan P Bollback, Thomas L York, and Rasmus Nielsen. Estimation of 2nes from temporal allele frequency data. *Genetics*, 179(1):497–502, May 2008.

[5] Alison F Feder, Sergey Kryazhimskiy, and Joshua B Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, February 2014.

[6] Jeremy J Berg and Graham Coop. A population genetic signal of polygenic adaptation. *PLoS Genet*, 10(8):e1004412, 2014.

[7] Matthieu Foll, Hyunjin Shim, and Jeffrey D Jensen. Wfabc: a w right–f isher abc-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular ecology resources*, 15(1):87–98, 2015.

[8] Joshua G Schraiber, Steven N Evans, and Montgomery Slatkin. Bayesian inference of natural selection from allele frequency time series. *Genetics*, 203(1):493–511, May 2016.

[9] Paula Tataru, Maria Simonsen, Thomas Bataillon, and Asger Hobolth. Statistical inference in the wright–fisher model using allele frequency data. *Systematic biology*, 66(1):e30–e46, 2017.

[10] Dennis Hedgecock. Does variance in reproductive success limit effective population sizes of marine organisms. *Genetics and evolution of aquatic organisms*, 122:122–134, 1994.

[11] Bjarki Eldon and John Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172(4):2621–2633, April 2006.

[12] Aurélien Tellier and Christophe Lemaire. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Mol. Ecol.*, 23(11):2637–2652, June 2014.

[13] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[14] Ramanan Laxminarayan, Brian Wahl, Shankar Reddy Dudala, K Gopal, S Neelima, KS Jawahar Reddy, J Radhakrishnan, Joseph A Lewnard, et al. Epidemiology and transmission dynamics of covid-19 in two indian states. *Science*, 370(6517):691–697, 2020.

[15] Dillon C Adam, Peng Wu, Jessica Y Wong, Eric HY Lau, Tim K Tsang, Simon Cauchemez, Gabriel M Leung, and Benjamin J Cowling. Clustering and superspreading potential of sars-cov-2 infections in hong kong. *Nature Medicine*, 26(11):1714–1719, 2020.

[16] Andrew M Sackman, Rebecca B Harris, and Jeffrey D Jensen. Inferring demography and selection in organisms characterized by skewed offspring distributions. *Genetics*, pages genetics–301684, 2019.

[17] S E Luria and M Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6):491–511, November 1943.

[18] Richard A Neher and Oskar Hallatschek. Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci. U. S. A.*, 110(2):437–442, January 2013.

[19] Oskar Hallatschek and David R Nelson. Gene surfing in expanding populations. *Theoretical population biology*, 73(1):158–170, 2008.

[20] Diana Fusco, Matti Gralka, Jona Kayser, Alex Anderson, and Oskar Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial luria–delbrück experiments. *Nature communications*, 7:12760, 2016.

[21] Gabriel Birzu, Oskar Hallatschek, and Kirill S Korolev. Fluctuations uncover a distinct class of traveling waves. *Proc. Natl. Acad. Sci. U. S. A.*, 115(16):E3645–E3654, April 2018.

[22] Gabriel Birzu, Oskar Hallatschek, and Kirill S Korolev. Genealogical structure changes as range expansions transition from pushed to pulled. *bioRxiv*, 2020.

[23] Jason Schweinsberg. Coalescent processes obtained from supercritical Galton–Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, July 2003.

[24] E Bolthausen and A S. Sznitman. On ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.*, 197(2):247–276, October 1998.

[25] Éric Brunet, Bernard Derrida, Alfred H Mueller, and Stéphane Munier. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Physical Review E*, 76(4):041104, 2007.

[26] Michael M Desai, Aleksandra M Walczak, and Daniel S Fisher. Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 193(2):565–585, 2013.

[27] Katya Kosheleva and Michael M Desai. The dynamics of genetic draft in rapidly adapting populations. *Genetics*, 195(3):1007–1025, 2013.

[28] Jason Schweinsberg et al. Rigorous results for a population model with selection ii: genealogy of the population. *Electronic Journal of Probability*, 22, 2017.

[29] Anne-Laure Basdevant, Christina Goldschmidt, et al. Asymptotics of the allele frequency spectrum associated with the bolthausen-sznitman coalescent. *Electronic Journal of Probability*, 13:486–512, 2008.

[30] Bjarki Eldon. Structured coalescent processes from a modified moran model with large offspring numbers. *Theoretical population biology*, 76(2):92–104, 2009.

[31] Bjarki Eldon. Estimation of parameters in large offspring number models and ratios of coalescence times. *Theoretical population biology*, 80(1):16–28, 2011.

[32] Oskar Hallatschek. Selection-like biases emerge in population models with recurrent jackpot events. *Genetics*, 210(3):1053–1073, 2018.

[33] Nicholas H Barton and Alison M Etheridge. The relation between reproductive value and genetic contribution. *Genetics*, 188(4):953–973, 2011.

[34] Chris Cannings. The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models. *Advances in Applied Probability*, pages 260–290, 1974.

[35] Jason Schweinsberg. Coalescent processes obtained from supercritical galton–watson processes. *Stochastic Processes and their Applications*, 106(1):107–139, 2003.

[36] Motoo Kimura et al. *Stochastic processes and distribution of gene frequencies under natural selection*. Citeseer, 1954.

[37] Crispin Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.

[38] M Kimura. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, 61(4):893–903, April 1969.

[39] Julien Berestycki, Nathanaël Berestycki, and Vlada Limic. Asymptotic sampling formulae for λ-coalescents. *Annales de l'IHP Probabilités et statistiques*, 50(3):715–731, 2014.

[40] Ricky Der and Joshua B Plotkin. The equilibrium allele frequency distribution for a population with reproductive skew. *Genetics*, 196(4):1199–1216, April 2014.

[41] Boris Vladimirovich Gnedenko and Andreĭ Kolmogorov. *Limit distributions for sums of independent random variables*, volume 233. Addison-wesley Reading, MA, 1968.

[42] Vladimir V Uchaikin and Vladimir M Zolotarev. *Chance and Stability: Stable Distributions and their Applications*. Walter de Gruyter, September 2011.

[43] Robert C Griffiths. The λ-fleming-viot process and a connection with wright-fisher diffusion. *Advances in Applied Probability*, 46(4):1009–1035, 2014.

[44] Alison M Etheridge, Robert C Griffiths, and Jesse E Taylor. A coalescent dual process in a moran model with genic selection, and the lambda coalescent limit. *Theoretical population biology*, 78(2):77–92, 2010.

[45] Samuel Karlin and Howard E Taylor. *A Second Course in Stochastic Processes*. Elsevier, June 1981.

[46] Warren J Ewens. The diffusion equation and a pseudo-distribution in genetics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2):405–412, 1963.

[47] Ivana Cvijović, Benjamin H Good, and Michael M Desai. The effect of strong purifying selection on genetic diversity. *Genetics*, 209(4):1235–1278, August 2018.

[48] Michael M Desai and Daniel S Fisher. Beneficial Mutation–Selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, July 2007.

[49] Daniel B Weissman, Michael M Desai, Daniel S Fisher, and Marcus W Feldman. The rate at which asexual populations cross fitness valleys. *Theoretical population biology*, 75(4):286–300, 2009.

[50] Richard A Neher and Boris I Shraiman. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188(4):975–996, 2011.

[51] Steven N Evans, Yelena Shvets, and Montgomery Slatkin. Non-equilibrium theory of the allele frequency spectrum. *Theoretical population biology*, 71(1):109–119, 2007.

[52] Ricky Der, Charles Epstein, and Joshua B Plotkin. Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics*, 191(4):1331–1344, August 2012.

Appendix A: Generalized central limit theorem

Here, we briefly summarize the generalized central limit theorem [41, 42]. Suppose that each random number u_i is sampled from the Pareto distribution $P(u) = \frac{\alpha}{u^{\alpha+1}}$ $(u \geq 1)$ and consider the shifted and rescaled random variable ζ;

$$\zeta = \frac{\sum_{i=1}^{n} u_i - a_n}{b_n}, \tag{A1}$$

where a_n and b_n are

$$\begin{aligned}
&a_n = 0, \quad b_n = \Big(\frac{\pi n}{2\Gamma(\alpha)\sin\frac{\pi\alpha}{2}}\Big)^{1/\alpha} \quad \text{for } 0 < \alpha < 1, \\
&a_n = n\log n, \quad b_n = \frac{\pi}{2}n \quad \text{for } \alpha = 1, \\
&a_n = \frac{\alpha}{\alpha-1}n, \quad b_n = \Big(\frac{\pi n}{2\Gamma(\alpha)\sin\frac{\pi\alpha}{2}}\Big)^{1/\alpha} \quad \text{for } 1 < \alpha < 2, \\
&a_n = \frac{\alpha}{\alpha-1}n = 2n, \quad b_n = (n\log n)^{1/2} \quad \text{for } \alpha = 2.
\end{aligned} \tag{A2}$$

It is well-known that the distribution of ζ is well-approximated by the α-stable distribution, which we denote as $P_\alpha(\zeta)$. While an explicit expression of $P_\alpha(\zeta)$ is not available in general, the characteristic function is given by

$$\begin{aligned}\langle e^{is\zeta}\rangle &= \int d\zeta e^{is\zeta} P_\alpha(\zeta) \\ &\sim \begin{cases} \exp\left[-|s|^\alpha\left(1+i\,\mathrm{sgn}(s)\frac{2}{\pi}\log|s|\right)\right] & (\alpha=1) \\ \exp\left[-|s|^\alpha\left(1-i\,\mathrm{sgn}(s)\tan\frac{\pi\alpha}{2}\right)\right] & (\alpha\neq 1)\end{cases}, \quad \text{for } n\to\infty. \end{aligned} \tag{A3}$$

Appendix B: The transition density of an allele frequency $w_N(y|x)$ and the asymptotic dynamics for lager N

Allele-frequency change in a generation is characterized by the transition density $w_N(y|x)$, which is the probability distribution of the allele frequency y at the next generation given the current allele frequency x. When N is lager, the asymptotic dynamics can be described by a time-continuous differential Chapman-Kolmogorov equation, which is defined by an advection velocity $V(x)$, diffusion coefficient $D(x)$, and jump kernel $w(y|x)$ [37]. The triplet is obtained from the transition density $w_N(y|x)$ as follows:

$$\begin{aligned} w(y|x) &= \lim_{N\to\infty}\frac{w_N(y|x)}{\delta t_N} \\ V(x) &= \lim_{N\to\infty}\frac{1}{\delta t_N}\int_{|y-x|<\epsilon}(y-x)w_N(y|x)dy \\ D(x) &= \lim_{N\to\infty}\frac{1}{\delta t_N}\int_{|y-x|<\epsilon}(y-x)^2 w_N(y|x)dy, \end{aligned} \tag{B1}$$

where δt_N is an N-dependent timescale, corresponding to one generation measured in units of the coalescent timescale. In the following, we derive the transition density $w_N(y|x)$ and the asymptotic dynamicsfor general α by using a similar computational technique used in [32], wherein the case of $\alpha=1$ is studied extensively.

As mentioned in the main text, when $\alpha\leq 2$, the binomial sampling error is negligible for large N compared to the stochasticity coming from broad offspring number fluctuations, and we can replace the binomial distribution in Equation 13 of the main text with the Dirac delta function;

$$w_N(y|x) = \langle\delta(y-\frac{M}{M+W})\rangle_{M,W} = \langle\int_{-\infty}^{+\infty}\frac{d\sigma}{2\pi}e^{i(y-\frac{M}{M+W})\sigma}\rangle_{M,W}. \tag{B2}$$

Here $\langle\cdot\rangle_{M,W}$ means the average over $M=\sum_{i=1}^{xN}u_i$ and $W=\sum_{i=1}^{(1-x)N}v_i$. Using the variable $s=\frac{\sigma}{M+W}$, we can rewrite w_N as

$$\begin{aligned} w_N(y|x) &= \langle\int_{-\infty}^{+\infty}\frac{ds}{2\pi}(M+W)e^{-is(M-y(M+W))}\rangle_{M,W} \\ &= \partial_y\int\frac{ds}{2\pi is}\langle e^{-isM(1-y)+isWy)}\rangle_{M,W} \\ &= \partial_y W_N(y|x). \end{aligned} \tag{B3}$$

Here,

$$W_N(y|x) = \int_{-\infty}^{+\infty}\frac{ds}{2\pi is}\Phi(-s(1-y);xN)\Phi(sy;(1-x)N) \tag{B4}$$

with

$$\Phi(s;n) = \langle e^{is\sum_{i=1}^n u_i}\rangle. \tag{B5}$$

To use the properties of the α-stable distributions in Appendix A, we further rewrite $W_N(y|x)$ as follows:

$$\begin{aligned}
W_N(y|x) &= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} \Phi(-s(1-y); xN)\Phi(sy; (1-x)N) \\
&= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} \langle e^{-is(1-y)\sum_{i=1}^{xN} x_i} \rangle \langle e^{isy\sum_{i=1}^{(1-x)N} x_i} \rangle \\
&= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} \left\langle e^{-is(1-y)(b_{xN}\zeta + a_{xN})} \right\rangle_{\zeta} \left\langle e^{isy(b_{(1-x)N}\zeta' + a_{(1-x)N})} \right\rangle_{\zeta'} \\
&= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-is(1-y)a_{xN} + isya_{(1-x)N}} \left\langle e^{-is(1-y)b_{xN}\zeta} \right\rangle_{\zeta} \left\langle e^{isyb_{(1-x)N}\zeta'} \right\rangle_{\zeta'}.
\end{aligned} \tag{B6}$$

When N is large, the quantities in the two brackets in the last line can be approximated by the characteristic functions of α-stable distribution, Equation A3, with $s \to -s(1-y)b_{xN}$ and $s \to syb_{(1-x)N}$, respectively. Thus, when $\alpha \neq 1$, Equation B6 can be computed as

$$\begin{aligned}
W_N(y|x) &= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-is(1-y)a_{xN} + isya_{(1-x)N}} \\
&\qquad \times e^{-|s|^\alpha (1-y)^\alpha b_{xN}^\alpha (1 + i\mathrm{sgn}(s)\tan\frac{\pi\alpha}{2})} e^{-|s|^\alpha y^\alpha b_{(1-x)N}^\alpha (1 - i\mathrm{sgn}(s)\tan\frac{\pi\alpha}{2})} \\
&= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-|s|^\alpha \{(1-y)^\alpha b_{xN}^\alpha + y^\alpha b_{(1-x)N}^\alpha\}} \\
&\qquad \times e^{-is(1-y)a_{xN} + isya_{(1-x)N}} e^{-i|s|^\alpha \mathrm{sgn}(s)\tan\frac{\pi\alpha}{2}\left((1-y)^\alpha b_{xN}^\alpha - y^\alpha b_{(1-x)N}^\alpha\right)} \\
&= \int_0^{\infty} \frac{ds}{\pi s} e^{-s^\alpha \{(1-y)^\alpha b_{xN}^\alpha + y^\alpha b_{(1-x)N}^\alpha\}} \\
&\quad \times \sin\left[-s\left((1-y)a_{xN} - ya_{(1-x)N}\right) - s^\alpha \tan\frac{\pi\alpha}{2}\left((1-y)^\alpha b_{xN}^\alpha - y^\alpha b_{(1-x)N}^\alpha\right)\right].
\end{aligned} \tag{B7}$$

In the following, we evaluate the integral expression of $W_N(y|x)$ and compute the transition density $w_N(y|x)$ from Equation B3.

When $\alpha < 1$

By using Equation A2,

$$a_n = 0, \quad b_n^\alpha = \pi(2\Gamma(\alpha)\sin\frac{\pi\alpha}{2})n \equiv c_\alpha n, \tag{B8}$$

we have

$$\begin{aligned}
W_N(y|x) &= \int_0^{\infty} \frac{ds}{\pi s} e^{-s^\alpha c_\alpha N((1-y)^\alpha x + y^\alpha(1-x))} \\
&\qquad \times \sin[-s^\alpha c_\alpha N \tan\frac{\pi\alpha}{2}((1-y)^\alpha x - y^\alpha(1-x))].
\end{aligned} \tag{B9}$$

By setting $Nc_\alpha s^\alpha = \sigma$, $W_N(y|x)$ becomes

$$\begin{aligned}
W_N(y|x) &= \frac{1}{\alpha}\int_0^{\infty} \frac{d\sigma}{\pi\sigma} e^{-\sigma((1-y)^\alpha x + y^\alpha(1-x))} \sin[-\sigma\tan\frac{\pi\alpha}{2}((1-y)^\alpha x - y^\alpha(1-x))] \\
&= -\frac{\tan^{-1}\left(\frac{\tan\left(\frac{\pi\alpha}{2}\right)(x(1-y)^\alpha - (1-x)y^\alpha)}{(1-x)y^\alpha + x(1-y)^\alpha}\right)}{\pi\alpha}.
\end{aligned} \tag{B10}$$

By differentiating it with respect to y, we obtain

$$w_N(y|x) = \frac{x(1-x)\sin(\pi\alpha)((1-y)y)^{\alpha-1}}{\pi\left(x^2(1-y)^{2\alpha} + (1-x)^2 y^{2\alpha} + 2x(1-x)\cos(\pi\alpha)((1-y)y)^\alpha\right)}. \tag{B11}$$

Note that this does not depend on N, which is consistent with the fact that the coalescent time is $\mathcal{O}(N^0)$ when $\alpha < 1$.

When $1 < \alpha < 2$

By using Equation A2,

$$a_n = \frac{\alpha}{\alpha - 1} n$$
$$b_n^\alpha = B_\alpha n, \text{ where } B_\alpha \equiv \frac{\pi}{2\Gamma(\alpha)\sin\frac{\pi\alpha}{2}}, \tag{B12}$$

Equation B7 becomes

$$W_N(y|x) = \int_0^\infty \frac{ds}{\pi s} e^{-B_\alpha N s^\alpha \{(1-y)^\alpha x + y^\alpha (1-x)\}} \times \sin[-s\frac{\alpha}{\alpha-1} N(x-y) - s^\alpha \tan\frac{\pi\alpha}{2} B_\alpha N\big((1-y)^\alpha x - y^\alpha(1-x)\big)]. \tag{B13}$$

By changing the variable of integration as $\sigma = N^{1/\alpha} s$, we have

$$W_N(y|x) = \int_0^\infty \frac{d\sigma}{\pi\sigma} e^{-B_\alpha \sigma^\alpha \{(1-y)^\alpha x + y^\alpha (1-x)\}} \times \sin[-\sigma\frac{\alpha}{\alpha-1} N^{1-\frac{1}{\alpha}}(x-y) - \sigma^\alpha \tan\frac{\pi\alpha}{2} B_\alpha\big((1-y)^\alpha x - y^\alpha(1-x)\big)]. \tag{B14}$$

By changing the variable of integration as $\sigma' = \frac{\alpha}{\alpha-1}|x-y|\sigma$ and redefining σ' as σ, we have

$$W_N(y|x) = \int_0^\infty \frac{d\sigma}{\pi\sigma} e^{-\mu_1\sigma^\alpha} \sin(-\mathrm{sgn}(x-y) N^{1-\frac{1}{\alpha}}\sigma - \mu_2\sigma^\alpha), \tag{B15}$$

where

$$\mu_1 = B_\alpha (\frac{\alpha-1}{\alpha})^\alpha \frac{(1-y)^\alpha x + y^\alpha(1-x)}{|x-y|^\alpha}$$
$$\mu_2 = \tan\frac{\pi\alpha}{2} B_\alpha (\frac{\alpha-1}{\alpha})^\alpha \frac{(1-y)^\alpha x - y^\alpha(1-x)}{|x-y|^\alpha}. \tag{B16}$$

The transition probability $w_N(y|x)$ is given by

$$w_N(y|x) = \partial_y W_N(y|x)$$
$$= \mathrm{sgn}(x-y)\frac{\partial_y\mu_1}{\pi}\int_0^\infty d\sigma\sigma^{\alpha-1} e^{-\mu_1\sigma^\alpha}\sin(N^{1-\frac{1}{\alpha}}\sigma + \mathrm{sgn}(x-y)\mu_2\sigma^\alpha)$$
$$-\frac{\partial_y\mu_2}{\pi}\int_0^\infty d\sigma\sigma^{\alpha-1} e^{-\mu_1\sigma^\alpha}\cos(N^{1-\frac{1}{\alpha}}\sigma + \mathrm{sgn}(x-y)\mu_2\sigma^\alpha). \tag{B17}$$

Consider the integral

$$J_\alpha = \int_0^\infty d\sigma\sigma^{\alpha-1} e^{-\mu\sigma^\alpha} e^{iN^{1-\frac{1}{\alpha}}\sigma} \tag{B18}$$

where $\mu = \mu_1 - i\,\mathrm{sgn}(x-y)\mu_2$. Then, the transition probability can be written as

$$w_N(y|x) = \mathrm{sgn}(x-y)\frac{\partial_y\mu_1}{\pi}\mathrm{Im}J_\alpha - \frac{\partial_y\mu_2}{\pi}\mathrm{Re}J_\alpha. \tag{B19}$$

From Watson's lemma, the integral J_α can be expressed as a series expansion;

$$J_\alpha = \sum_{m=1}^\infty \frac{1}{N^{m(\alpha-1)}} e^{i\frac{\pi}{2}m\alpha}(-\mu)^{m-1}\frac{\Gamma(m\alpha)}{\Gamma(m)}. \tag{B20}$$

By substituting Equation B20 into Equation B19 and writing $\mu = |\mu|e^{i\theta}$, we obtain

$$w_N(y|x) = \sum_{m=1}^{\infty} \frac{(-|\mu|)^{m-1}}{N^{m(\alpha-1)}} \frac{\Gamma(m\alpha)}{\pi\Gamma(m)} \left[\mathrm{sgn}(x-y)\partial_y\mu_1 \sin(\frac{\pi}{2}m\alpha + (m-1)\theta) - \partial_y\mu_2 \cos(\frac{\pi}{2}m\alpha + (m-1)\theta)\right] \tag{B21}$$

The leading order ($m = 1$) is given by

$$\begin{aligned} w_N(y|x) &= \frac{\Gamma(\alpha)}{N^{\alpha-1}} \left(\mathrm{sgn}(x-y)\frac{\partial_y\mu_1}{\pi}\sin\frac{\pi\alpha}{2} - \frac{\partial_y\mu_2}{\pi}\cos\frac{\pi\alpha}{2}\right) \\ &= \begin{cases} N^{1-\alpha}\alpha\left(\frac{\alpha-1}{\alpha}\right)^{\alpha} x(1-x)\frac{(1-y)^{\alpha-1}}{(y-x)^{\alpha+1}} & \text{when } x < y \\ N^{1-\alpha}\alpha\left(\frac{\alpha-1}{\alpha}\right)^{\alpha} x(1-x)\frac{y^{\alpha-1}}{(x-y)^{\alpha+1}} & \text{when } x > y. \end{cases} \end{aligned} \tag{B22}$$

Equation 21 in the main text can be obtained by introducing the continuous time $\tau \equiv t/(C_\alpha N^{\alpha-1})$ where $C_\alpha \equiv \alpha\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}$. Equation 22 follows from the neutrality $\frac{d}{dt}\langle x\rangle = 0$. Note that the expansion of Equation B20 is possible only when $|x-y|$ is finite, i.e., when $|x-y| > \epsilon$ where ϵ is an N-independent positive constant. Although $w_N(y|x)$ in B22 diverges as $|x-y| \to 0$, this divergence is not a problem, because the jump term of the asymptotic dynamics in Equation 20 can be obtained from $w_N(y|x)$ for $|x-y| > \epsilon$ (see [37]).

When $\alpha = 2$

a_n and b_n are given by

$$a_n = \frac{\alpha}{\alpha-1}n = 2n, \quad b_n = (n\log n)^{1/2}. \tag{B23}$$

Equation B7 then becomes

$$\begin{aligned} W_N(y|x) &= \int_0^\infty \frac{ds}{\pi s} e^{-s^2\{(1-y)^2 xN\log xN + y^2(1-x)N\log(1-x)N\}} \times \sin\big(-2sN(x-y)\big) \\ &= \int_0^\infty \frac{ds}{\pi s} e^{-s^2\left\{(-2xy+x+y^2)N\log N + \left((1-y)^2 x\log x + y^2(1-x)\log(1-x)\right)N\right\}} \times \sin\big(-2sN(x-y)\big). \end{aligned}$$

By changing the variable of integration as $\sigma = (N\log N)^{\frac{1}{2}}s$,

$$\begin{aligned} W_N(y|x) &= \int_0^\infty \frac{d\sigma}{\pi\sigma} e^{-\sigma^2\left\{(-2xy+x+y^2) + \left((1-y)^2 x\log x + y^2(1-x)\log(1-x)\right)(\log N)^{-1}\right\}} \times \sin\big(-2(\frac{N}{\log N})^{\frac{1}{2}}\sigma(x-y)\big) \\ &\approx \int_0^\infty \frac{d\sigma}{\pi\sigma} e^{-\sigma^2\ (-2xy+x+y^2)} \times \sin\big(-2(\frac{N}{\log N})^{\frac{1}{2}}\sigma(x-y)\big) \\ &= -\frac{1}{2}\mathrm{erf}\left((x-y)\sqrt{\frac{N}{\log(N)\,(-2xy+x+y^2)}}\right), \end{aligned} \tag{B24}$$

where $\mathrm{erf}(x)$ is the Gauss error function

$$\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}dt. \tag{B25}$$

By differentiating $W_N(y|x)$ with respect to y, we have

$$w_N(y|x) = (\frac{N}{\log N})^{\frac{1}{2}} \frac{1}{\sqrt{\pi}} \frac{(1-x)x}{(-2xy+x+y^2)^{3/2}} e^{-\frac{N(x-y)^2}{\log N(-2xy+x+y^2)}}. \tag{B26}$$

Suppose that ϵ is a sufficiently small but finite constant. For $|x-y|<\epsilon$, w_N can be approximated as

$$\begin{aligned} w_N(y|x) &= (\frac{N}{\log N})^{\frac{1}{2}} \frac{1}{\sqrt{\pi}} \frac{1}{(x(1-x))^{1/2}} e^{-\frac{N(x-y)^2}{\log N(x(1-x))}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-y)^2}{2\sigma^2}}. \end{aligned} \tag{B27}$$

where $2\sigma^2 = \frac{\log N}{N} x(1-x)$. From the symmetry $y-x \to -(y-x)$ of $w_N(y|x)$, the advection term is zero. The diffusivity D is given by

$$D = \frac{1}{\delta t_N} \int_{|x-y|<\epsilon} dy\ (x-y)^2 w_N(y|x) = \frac{1}{\delta t_N}\sigma^2 = \frac{1}{\delta t_N}\frac{\log N}{N}\frac{1}{2}x(1-x) = \frac{1}{2}x(1-x), \tag{B28}$$

where we have introduced the natural timescale as $\delta t_N = \frac{\log N}{N}$ and used the integral approximation

$$\int_{-\epsilon}^{\epsilon} d\Delta\ \Delta^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\Delta^2}{2\sigma^2}) = \frac{1}{\sqrt{2\pi}}\sigma\left(\sqrt{2\pi}\sigma \mathrm{erf}\left(\frac{\epsilon}{\sqrt{2}\sigma}\right) - 2\epsilon e^{-\frac{\epsilon^2}{2\sigma^2}}\right) \approx \sigma^2. \tag{B29}$$

Finally, the jump kernel asymptotically vanishes on the time scale δt_N,

$$w(y|x) = \lim_{N\to\infty} \frac{w_N(y|x)}{\delta t_N}, \tag{B30}$$

because for fixed x, y with $|x-y|>\epsilon$, $w_N(y|x)$ becomes exponentially small as N becomes large.

Thus, in the large-N limit, $\alpha = 2$ corresponds to Wright-Fisher diffusion for a population of effective size

$$N_e = N\log(N). \tag{B31}$$

When $\alpha > 2$

In this case, since the Pareto distribution $P(u) = \frac{\alpha}{u^{\alpha+1}}$ $(u \geq 1)$ has finite mean $a = \frac{\alpha}{\alpha-1}$ and finite variance $b^2 = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$, and the large N limit of the allele frequency dynamics should be described by the Wright-Fisher diffusion process. To confirm this more generally, we consider a general distribution with finite mean and variance, namely, consider that each individual's offspring number u_i is sampled from a distribution with mean a and variance b^2. Then, from the central limit theorem, the shifted and rescaled variable

$$\zeta = \frac{\sum_{i=1}^n x_i - a_n}{b_n}, \text{where } a_n = an,\ b_n = \sqrt{n}b, \tag{B32}$$

obeys the normal distribution $\mathcal{N}(0,1)$. Its characteristic function is given by $f(s) = \exp(-\frac{1}{2}s^2)$. Thus, we have

$$\begin{aligned} W_N(y|x) &\approx \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-is(1-y)a_{xN} + isya_{(1-x)N}} f(-s(1-y)b_{xN}) f(syb_{(1-x)N}) \\ &= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-is(1-y)axN + isya(1-x)N} f(-s(1-y)\sqrt{xN}b) f(sy\sqrt{(1-x)N}b) \\ &= \int_{-\infty}^{+\infty} \frac{ds}{2\pi i s} e^{-is(1-y)axN + isya(1-x)N} e^{-\frac{1}{2}s^2(1-y)^2 xNb^2} e^{-\frac{1}{2}s^2y^2(1-x)Nb^2}. \end{aligned} \tag{B33}$$

By setting $\sigma = N^{1/2}s$,

$$\begin{aligned} W_N(y|x) &= \int_{-\infty}^{+\infty} \frac{d\sigma}{2\pi i\sigma} e^{-ia(x-y)N^{1/2}\sigma} e^{-\frac{1}{2}b^2\sigma^2((1-y)^2x + y^2(1-x))} \\ &= \int_0^{+\infty} \frac{d\sigma}{\pi\sigma} \sin(-a(x-y)N^{1/2}\sigma) e^{-\frac{1}{2}b^2\sigma^2((1-y)^2x + y^2(1-x))} \\ &= -\frac{1}{2}\mathrm{erf}\left(\frac{a\sqrt{N}(x-y)}{\sqrt{2}b\sqrt{(-2xy + x + y^2)}}\right). \end{aligned} \tag{B34}$$

Thus, we obtain

$$w_N(y|x) = \partial_y W_N(y|x) = \sqrt{N}\sqrt{\frac{1}{2\pi}}\gamma x(1-x)\frac{\exp\left(-\frac{\gamma^2 N(x-y)^2}{2(-2xy+x+y^2)}\right)}{(-2xy+x+y^2)^{3/2}}. \tag{B35}$$

where $\gamma \equiv a/b$. For the Pareto distribution, $\gamma = (\frac{\alpha}{\alpha-1})/\sqrt{\frac{\alpha}{(\alpha-1)^2(\alpha-2)}} = \sqrt{\alpha(\alpha-2)}$.

For $|x-y| > \epsilon$, $w_N(y|x)$ becomes exponentially small as N becomes large, and so the jump term does not exist in the asymptotic dynamics; $w(y|x) = 0$. For $|x-y| < \epsilon$, we can approximate $w_N(y|x)$ as

$$\begin{aligned} w_N(y|x) &= \sqrt{\frac{N\gamma^2}{2\pi x(1-x)}}\exp\left(-\frac{\gamma^2 N(x-y)^2}{2x(1-x)}\right) \\ &= \frac{1}{\sqrt{2\pi\Sigma^2}}e^{-\frac{(x-y)^2}{2\Sigma^2}}, \end{aligned} \tag{B36}$$

where $\Sigma^2 = \frac{x(1-x)}{\gamma^2 N}$. From the symmetry $y-x \to -(y-x)$ of $w_N(y|x)$, the advection is zero. Finally, the diffusion is evaluated as

$$\int_{|x-y|<\epsilon} dy(x-y)^2 w_N(y|x) = \Sigma\left(\Sigma\ \mathrm{erf}\left(\frac{\epsilon}{\sqrt{2}\Sigma}\right) - \sqrt{\frac{2}{\pi}}\epsilon e^{-\frac{\epsilon^2}{2\Sigma^2}}\right) \approx \Sigma^2 = \frac{x(1-x)}{\gamma^2 N}. \tag{B37}$$

Thus, by re-scaling time as $\tau = \frac{t}{\gamma^2 N}$, we obtain

$$D = x(1-x), \tag{B38}$$

which corresponds to the Wright-Fisher diffusion of a population of effective size $N_e = N\gamma^2 = N\alpha(\alpha-2)$. Notice $N_e \to 0$ as $\alpha \to 2$, indicating that the concept of the effective population size breaks down when the variance of the offspring distribution diverges.

Appendix C: The transition density for the differential Chapman-Kolmogorov equation for $1 < \alpha < 2$

Here we derive the short-time transition density given in Equations 23 and 24 and determine $g(\xi)$ in the scaling ansatz given in Equation 39.

1. The short-time transition density

Before discussing the CK equation in Equation 20, it is instructive to start from the simple diffusion equation,

$$\partial_\tau P(x,\tau) = D\partial_x^2 P(x,\tau), \tag{C1}$$

with the initial condition $P(x,\tau=0) = \delta(x-x_0)$. The solution of this initial value problem is given by

$$P(\Delta x,\tau) = \frac{1}{\sqrt{2\pi(2D\tau)}}\exp\left(-\frac{\Delta x^2}{2(2D\tau)}\right), \tag{C2}$$

which is usually derived from the Laplace-Fourier transformation. However, this solution can also be obtained by using the central limit theorem: Equation C1 is equivalent to a Brownian motion where jumps $X \to X \pm a$ occur with rate $\frac{m}{2}$, where a and m are related with D via $D = \frac{a^2 m}{2}$. Since $n \approx m\tau$ jumps occur in time τ, the displacement is approximately given by $\Delta X(\tau) \approx \sum_{i=1}^n l_i$ where $l_i = \pm a$. Then, from the central limit theorem, $\Delta X(\tau)$ is distributed according to the normal distribution with mean $n\langle l_i\rangle = 0$ and variance $n\langle l_i^2\rangle = (m\tau)a^2 = 2D\tau$, namely, Equation C2. Note that, even if the diffusion constant depends on x, the solution in Equation C2 (with $D \to D(x_0)$) is valid in short times.

Essentially the same argument can be applied to the CK dynamics, except that the generalized central limit theorem should be employed since the variance of jump sizes is divergent in the case of the CK dynamics. Suppose that the initial density is given by $P(x',\tau=0) = \delta(x'-x)$ (for notational simplicity, the subscript 0 on x is dropped). In the

CK dynamics, the frequency change $\Delta X(\tau) = X(\tau) - x$ is caused by the bias $V(x)$ in Equation 22 and by stochastic jumps. The rate of a frequency-increasing jump and that of a frequency-decreasing jump are given by

$$W_+(x) = \int_{x+\epsilon}^{1} w(x'|x)dx' = \frac{x}{\alpha}\left(\frac{1-x}{\epsilon}\right)^{\alpha}, \tag{C3}$$

$$W_-(x) = \int_{0}^{x-\epsilon} w(x'|x)dx' = \frac{1-x}{\alpha}\left(\frac{x}{\epsilon}\right)^{\alpha}, \tag{C4}$$

respectively. Therefore, the expected number n of jump events in time τ is given by

$$n = (W_- + W_+)\tau. \tag{C5}$$

Because randomness in the number of jump events is negligible compared to that in jump sizes, it can be assumed that exactly n jumps occur in time τ. Then, the displacement $\Delta X(\tau) = X(\tau) - x$ can be written as

$$\Delta X(\tau) = V(x)\tau + \sum_{i=1}^{n} l_i, \tag{C6}$$

where $l_i \in [-x, -\epsilon] \cup [\epsilon, 1-x]$ denotes the displacement due to the i-th jump. For small τ, $w(y|x(\tau')) \approx w(y|x)$ for $0 < \tau' < \tau$, which means that $l_1, \ldots, l_n$ are independent and identically distributed. From Equation 21, each l_i is approximately sampled from the following power-law distribution,

$$P(l) = \begin{cases} \frac{W_+}{W_+ + W_-}\frac{\epsilon^\alpha \alpha}{l^{\alpha+1}} & (l \in [+\epsilon, +\infty)) \\ 0 & (l \in (-\epsilon, +\epsilon)) \\ \frac{W_-}{W_+ + W_-}\frac{\epsilon^\alpha \alpha}{|l|^{\alpha+1}} & (l \in (-\infty, -\epsilon]) \end{cases}, \tag{C7}$$

where the factor $\frac{W_+}{W_- + W_+}$ (resp. $\frac{W_-}{W_- + W_+}$) represents the probability that a given jump is frequency-increasing (resp. frequency-decreasing). $P(l)$ is normalized as $\int_{-\infty}^{\infty} P(l)dl = 1$. Note that, in Equation C7, the original range $[-x, -\epsilon] \cup [\epsilon, 1-x]$ of l has been extended to $[(-\infty, -\epsilon] \cup [\epsilon, \infty)$. Under this modification, the variance $\langle x(\tau)^2 \rangle$ is no longer well-defined. However, this modification does not alter short-time properties of typical events, because the presence of the boundaries at $x = 0, 1$ is not important for them.

By noting that $P(l)$ has a divergent variance and that the number of jumps is $n \approx \frac{\tau}{\epsilon^\alpha} \gg 1$ even for small τ (as $\epsilon \to +0$), the generalized central limit theorem states that the sum $\sum_{i=1}^{n} l_i$ in Equation C6 obeys an α-stable distribution. The stable distribution is characterized by $\langle l \rangle, \beta, \gamma$ given below (see, for example, [42]): The mean $\langle l \rangle$ is

$$\langle l \rangle = \frac{W_+ - W_-}{W_+ + W_-}\frac{\alpha}{\alpha - 1}\epsilon = \frac{x(1-x)^\alpha - x^\alpha(1-x)}{x^\alpha(1-x) + x(1-x)^\alpha}\frac{\alpha}{\alpha-1}\epsilon. \tag{C8}$$

Asymptotically, $P(l)$ satisfies

$$\begin{aligned} \int_{l}^{\infty} P(l')dl' &= \frac{W_+}{W_- + W_+}\frac{\epsilon^\alpha}{l^\alpha} \equiv \frac{c_+}{l^\alpha} \quad (l \to \infty), \\ \int_{-\infty}^{l} P(l')dl' &= \frac{W_-}{W_- + W_+}\frac{\epsilon^\alpha}{|l|^\alpha} \equiv \frac{c_-}{|l|^\alpha} \quad (l \to -\infty). \end{aligned} \tag{C9}$$

Note $c_- + c_+ = \epsilon^\alpha$. The parameters γ and β are determined from $c_\pm$;

$$\gamma \equiv \left(\frac{\pi(c_+ + c_-)n}{2\Gamma(\alpha)\sin\frac{\pi\alpha}{2}}\right)^{\frac{1}{\alpha}} = \epsilon\left(\frac{\pi n}{2\Gamma(\alpha)\sin\frac{\pi\alpha}{2}}\right)^{\frac{1}{\alpha}} = \left(\tau\frac{\pi(x(1-x)^\alpha + (1-x)x^\alpha)}{2\Gamma(\alpha+1)\sin\frac{\pi\alpha}{2}}\right)^{\frac{1}{\alpha}} \tag{C10}$$

$$\beta \equiv \frac{c_+ - c_-}{c_+ + c_-} = \frac{W_+ - W_-}{W_+ + W_-} = \frac{x(1-x)^\alpha - x^\alpha(1-x)}{x^\alpha(1-x) + x(1-x)^\alpha}. \tag{C11}$$

Then, from the generalized central limit theorem, the random variable,

$$Z \equiv \frac{\sum_{i=1}^{n} l_i - n\langle l \rangle}{\gamma}, \tag{C12}$$

has the following characteristic function,

$$\langle e^{ikZ}\rangle = \int e^{ikz} P_Z(z)dz \overset{\epsilon\to+0}{=} \exp\left[-|k|^\alpha\left(1 - i\beta\tan\frac{\pi\alpha}{2}\mathrm{sign}k\right)\right]. \tag{C13}$$

We can determine the characteristic function for Δx, using Equation C13 and the relation

$$\Delta X(\tau) = V(x)\tau + \gamma Z + n\langle l\rangle, \tag{C14}$$

which follows from Equations C6 and C12. While $V(x)$ and $\langle l\rangle$ are divergent in the limit $\epsilon \to +0$, we can show, by using Equation C8 and $V(x) = -\int_{|x-x'|>\epsilon} dx'(x'-x)w(x'|x) \approx \frac{1}{\epsilon^{\alpha-1}}\frac{1}{\alpha-1}\left(x^\alpha(1-x) - x(1-x)^\alpha\right)$, that these divergent terms exactly cancel out each other. Therefore, the displacement is simplified as

$$\Delta X(\tau) = \gamma Z. \tag{C15}$$

Equations 24 and 23 in the main text are the same as Equations C15 and C13 (with the replacement of $x \to x_0$). By substituting this into Equation C13, we obtain the characteristic function of the allele frequency $X(\tau)$;

$$\langle e^{ikX(\tau)}\rangle = \int e^{ikx'} P(x',\tau|x_0)dx' = \exp\left[ikx_0 - |\gamma(x_0)k|^\alpha\left(1 - i\beta(x_0)\tan\frac{\pi\alpha}{2}\mathrm{sign}k\right)\right]. \tag{C16}$$

2. The scaling ansatz for the long-time transition density in Equation 39

Consider the initial distribution $P(x,\tau=0) = \delta(x-x_0)$ with $x_0 \ll 1$. After some time, the distribution spreads over the region $x \ll 1$ with a peak at the extinction boundary $x = 0$. As presented in Equation 39 of the main text, up to a constant prefactor, $P(x,\tau)$ takes the following form

$$P(x,\tau) \sim \tau^{-2\eta} g(\xi),$$

where $\eta = (\alpha-1)^{-1}$ and $\xi = \frac{x}{\tau^\eta}$. Here, we present an analytic argument to determine $g(\xi)$.

Equation 20 can be rewritten as

$$\frac{\partial P}{\partial \tau} = \int_{|\Delta|<\epsilon} d\Delta(f_\Delta(x-\Delta)P(x-\Delta,\tau) - f_\Delta(x)P(x,\tau)) + \frac{\partial}{\partial x}\int_{|\Delta|<\epsilon} d\Delta(f_\Delta(x)P(x,\tau)), \tag{C17}$$

where $f_\Delta(x) \equiv w(x+\Delta|x)$ given by Equation 21. For $x \ll 1$, $f_\Delta(x)$ is approximately given by

$$f_\Delta(x) = \begin{cases} \frac{x}{\Delta^{\alpha+1}} & (\Delta > 0) \\ \frac{x(x+\Delta)^{\alpha-1}}{\Delta^{\alpha+1}} & (\Delta < 0) \end{cases}. \tag{C18}$$

We substitute the ansatz $P(x,\tau) \sim \tau^{-2\eta}g(\xi)$ into the above CK equation. The left-hand side of the CK equation becomes

$$\mathrm{LHS} = -2\eta\tau^{-2\eta-1}g(\xi) - \eta\tau^{-2\eta-1}g'(\xi)\xi, \tag{C19}$$

which is proportional to $\tau^{-2\eta-1} = \tau^{-\frac{\alpha+1}{\alpha-1}}$. The right-hand side is decomposed into the integrals over $\Delta > 0$ and those over $\Delta < 0$. We can show that the former is proportional to $\tau^{-\frac{\alpha+1}{\alpha-1}}$, while the latter is proportional to $\tau^{-\frac{2}{\alpha-1}}$; For example, one of the integrals over $\Delta > 0$ is

$$\int_{\Delta>0} d\Delta f_\Delta(x)P(x) = \int_{\Delta>0} d\Delta\frac{x}{\Delta^{\alpha+1}}\tau^{-2\eta}g(\xi) = \tau^{-\frac{\alpha+1}{\alpha-1}}\int_{\delta>0} d\delta\frac{\xi}{\delta^{\alpha+1}}g(\xi),$$

while one of the integrals over $\Delta < 0$ is

$$\int_{\Delta<0} d\Delta f_\Delta(x)P(x) = \int_{\Delta<0} d\Delta\frac{x(x+\Delta)^{\alpha-1}}{\Delta^{\alpha+1}}\tau^{-2\eta}g(\xi) = \tau^{-\frac{2}{\alpha-1}}\int_{\delta>0} d\delta\frac{\xi}{\delta^{\alpha+1}}g(\xi),$$

where we have changed the integration variable from Δ to $\delta = \frac{\Delta}{\tau^\eta}$. Since the extinction time for the initial frequency $x_0 \ll 1$ is much shorter than the coalescent timescale, we can assume $\tau \ll 1$, which implies that the integrals over

$\Delta > 0$ are negligible compared to those over $\Delta > 0$. By evaluating the integrals over $\Delta > 0$ using the scaling form of $P(x,\tau)$ and comparing them with Equation C19, we have

$$-\eta\Big(2g(\xi)+\xi g'(\xi)\Big) = \int_0^\infty \frac{d\delta}{\delta^{\alpha+1}}\Big((\xi-\delta)g(\xi-\delta)\Theta(\xi-\delta)-\xi g(\xi)+\delta\frac{d}{d\xi}(\xi g(\xi))\Big), \tag{C20}$$

where $\Theta(\cdot)$ is the Heaviside step function. Note that the variable of integration has been changed from Δ to $\delta = \frac{\Delta}{\tau^\eta}$, and the upper bound in the integral has been extended into $+\infty$, to make the equation analytically tractable. It is convenient to express Equation C20 in terms of $\Phi(\xi) \equiv \xi g(\xi)$;

$$-\eta\Big(\frac{\Phi(\xi)}{\xi}+\Phi'(\xi)\Big) = \int_0^\infty \frac{d\delta}{\delta^{\alpha+1}}\Big(\Phi(\xi-\delta)\Theta(\xi-\delta)-\Phi(\xi)+\delta\Phi'(\xi)\Big). \tag{C21}$$

The solution of the integro-differential equation in Equation C21 can be obtained as a series expansion. Assume, for small ξ,

$$\Phi(\xi) = c_1\xi^\beta + \cdots, \tag{C22}$$

where c_1 is a normalization and the exponent of the leading term is denoted by $\beta \in (0,1)$. Here, $\beta < 1$ is required since we are considering the situation where $P(x,\tau)$ is monotonically decreasing in x, while $\beta > 0$ is required to make $P(x,\tau)$ normalizable. By substituting Equation C22 into Equation C21, we have

$$-\frac{\beta+1}{\alpha-1}\frac{1}{\xi^{1-\beta}} + \cdots = \frac{\Gamma(-\alpha)\Gamma(1+\beta)}{\Gamma(1-\alpha+\beta)}\frac{1}{\xi^{\alpha-\beta}} + \cdots. \tag{C23}$$

Since $\frac{1}{\xi^{1-\beta}} \ll \frac{1}{\xi^{\alpha-\beta}}$ for $\xi \ll 1$, in order for the two sides to be balanced, the coefficient $\frac{\Gamma(-\alpha)\Gamma(1+\beta)}{\Gamma(1-\alpha+\beta)}$ needs to be zero, which is possible only when $\Gamma(1-\alpha+\beta)$ diverges. Since $1 < \alpha < 2$ and $0 < \beta < 1$, we can conclude $\beta = \alpha - 1$. Therefore, the leading term of $g(\xi)$ is given by

$$g(\xi) = \frac{c_1}{\xi^{2-\alpha}} + \cdots \quad (\xi \ll 1). \tag{C24}$$

More generally, by starting from the ansatz,

$$\Phi(\xi) = \sum_{m=1}^\infty c_m \xi^{(\alpha-1)m}, \tag{C25}$$

the coefficients $c_2, c_3, \cdots$ can be determined iteratively:

$$c_{m+1} = -\frac{1+(\alpha-1)m}{\alpha-1}\frac{\Gamma(m(\alpha-1))}{\Gamma(-\alpha)\Gamma(m(\alpha-1)+\alpha)}c_m \quad (m = 1,2,\cdots). \tag{C26}$$

By using this iteratively, we can express $\Phi(\xi)$ as

$$\Phi(\xi) = c_1\sum_{m=1}^\infty (-1)^{m+1}\frac{\Gamma(\alpha+1)\left(\frac{\alpha}{\alpha-1}\right)_{m-1}}{\alpha(\alpha-1)^m\Gamma(-\alpha)^{m-1}\Gamma(m+1)\Gamma(m(\alpha-1))}\xi^{(\alpha-1)m}, \tag{C27}$$

where $(\frac{\alpha}{\alpha-1})_{m-1}$ is the Pochhammer symbol, $(q)_n = \Gamma(q+n)/\Gamma(q)$. The analytic expression of $g(\xi)$ can be obtained from this using $g(\xi) = \frac{\Phi(\xi)}{\xi}$.

On the other hand, for $\xi \gg 1$, we expect that $g(\xi)$ decreases in the same way as the offspring distribution does;

$$g(\xi) \sim \frac{1}{\xi^{\alpha+1}} + \cdots \quad (\xi \gg 1). \tag{C28}$$

Therefore, we expect there is a crossover point ξ_c such that $g(\xi) \sim \frac{1}{\xi^{2-\alpha}} + \cdots$ for $\xi \ll \xi_\mathrm{c}$ and $g(\xi) \sim \frac{1}{\xi^{\alpha+1}} + \cdots$ for $\xi \gg \xi_\mathrm{c}$. The scaling form for $\xi \gg \xi_\mathrm{c}$ can indeed be confirmed by considering the following ansatz for $\Phi(\xi)$,

$$\Phi(\xi) = \begin{cases} c_1\xi^{\alpha-1} + \cdots & (\xi < \xi_\mathrm{c}) \\ c'\xi^{-\alpha'} + \cdots & (\xi > \xi_\mathrm{c}) \end{cases}, \tag{C29}$$

where c' is a normalization and α' is an exponent to be determined. Substituting this ansatz into Equation C21, we can show $\alpha' = \alpha$, leading to $g(\xi) \sim \frac{1}{\xi^{\alpha+1}} + \cdots$ for $\xi > \xi_c$.

Finally, we remark that, while Equation C27 is derived assuming $\xi \ll 1$, the series converges for any $\xi > 0$. This indicates that the scaling form $g(\xi) \sim \frac{1}{\xi^{\alpha+1}} + \cdots$ for large ξ should directly follow from a resummation of the infinite series in Equation C27. In fact, numerical evaluation of a finite truncation of the series indicates the crossover behavior Equation C29 (see Figure 15).

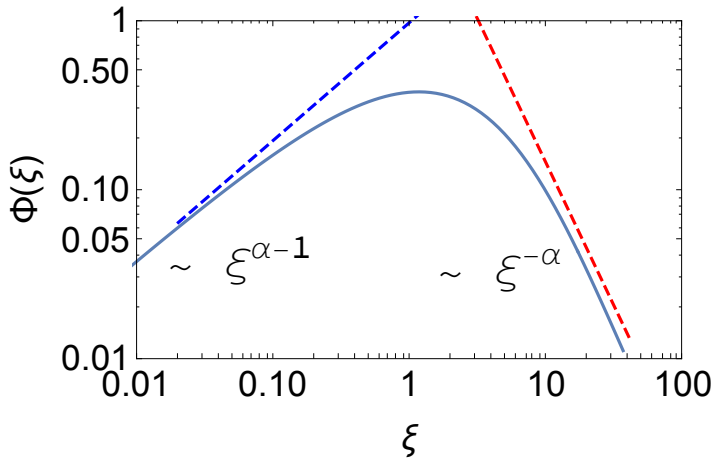

Figure 15. The infinite series in Equation C27 is evaluated numerically by truncating at $m = 150$ and using the van Wijngaarden transformation (solid line). $\alpha = 1.7$ is used. The dashed blue and red lines represent the asymptotic behaviors given in Equation C29.

Appendix D: From Lambda-Fleming-Viot Generator to differential Chapman-Kolmogorov equation

In Appendix B, the jump density $w(y|x)$ is derived from the generalized Wright-Fisher sampling, Equation 13 in the main text. Here, we present another more formal derivation of the jump density $w(y|x)$ for $1 < \alpha < 2$. See [32] for the case $\alpha = 1$.

1. Jump density for general Λ measure

The backward generator of the Λ coalescent process for the biallelic model (see e.g. [43, 44]) is given by

$$\mathcal{L}G_\tau(x|x_0) = \int_0^1 \Big(x_0 G_\tau(x|x_0 + (1-x_0)\lambda) - G_\tau(x|x_0) + (1-x_0) G_\tau(x|x_0 - x_0\lambda) \Big) \frac{\Lambda(d\lambda)}{\lambda^2}. \tag{D1}$$

This can be rewritten as a sum of two terms:

$$\mathcal{L}G_\tau(x|x_0) = A + B, \tag{D2}$$

where

$$A = x_0 \int_0^1 \Big(G_\tau(x|x_0 + (1-x_0)\lambda) - G_\tau(x|x_0) - (1-x_0)\lambda \partial_{x_0} G_\tau(x|x_0) \Big) \frac{\Lambda(d\lambda)}{\lambda^2}, \tag{D3}$$

$$B = (1-x_0) \int_0^1 \Big(G_\tau(x|x_0 - x_0\lambda) - G_\tau(x|x_0) + x_0\lambda \partial_{x_0} G_\tau(x|x_0) \Big) \frac{\Lambda(d\lambda)}{\lambda^2}. \tag{D4}$$

We introduce the integration variable $x' \equiv x_0 + (1-x_0)\lambda$ for A and $x' \equiv x_0 - x_0\lambda$ for B respectively. By writing

$$\frac{\Lambda(d\lambda)}{\lambda^2} = \frac{l(\lambda)}{\lambda^2} d\lambda, \tag{D5}$$

A and B become

$$A = x_0(1-x_0)\int_{x_0}^{1}\Big(G_\tau(x|x') - G_\tau(x|x_0) - (x'-x_0)\partial_{x_0}G_\tau(x|x_0)\Big)\frac{l(\frac{x'-x_0}{1-x_0})}{(x'-x_0)^2}dx', \tag{D6}$$

$$B = x_0(1-x_0)\int_{0}^{x_0}\Big(G_\tau(x|x') - G_\tau(x|x_0) - (x'-x_0)\partial_{x_0}G_\tau(x|x_0)\Big)\frac{l(\frac{x_0-x'}{x_0})}{(x'-x_0)^2}dx'. \tag{D7}$$

Defining the jump kernel $w(x|x_0)$ as

$$w(x'|x_0) = \begin{cases} \frac{x_0(1-x_0)}{(x'-x_0)^2}l(\frac{x'-x_0}{1-x_0}) & (x' > x_0), \\ \frac{x_0(1-x_0)}{(x'-x_0)^2}l(\frac{x_0-x'}{x_0}) & (x' < x_0), \end{cases} \tag{D8}$$

we can formally rewrite the generator as

$$\mathcal{L}G_\tau(x|x_0) = V(x_0)\partial_{x_0}G_\tau(x|x_0) + \mathrm{PV}\int_0^1 w(x'|x_0)[G_\tau(x|x') - G_\tau(x|x_0)]dx', \tag{D9}$$

where

$$V(x_0) = -\mathrm{PV}\int_0^1 dx' w(x'|x_0)(x'-x_0). \tag{D10}$$

2. When the measure is the Beta distribution Beta$(\alpha, 2-\alpha)$:

We take the Beta$(\alpha, 2-\alpha)$ distribution as the Λ measure, which corresponds to the descendant distribution considered in this study, $\sim 1/u^{1+\alpha}$:

$$\frac{\Lambda(d\lambda)}{\lambda^2} = \frac{l(\lambda)d\lambda}{\lambda^2} = \frac{\lambda^{+1-\alpha}(1-\lambda)^{\alpha-1}}{B(\alpha,2-\alpha)}\frac{d\lambda}{\lambda^2} = \frac{\lambda^{-1-\alpha}(1-\lambda)^{\alpha-1}}{B(\alpha,2-\alpha)}d\lambda. \tag{D11}$$

With this measure, A and B become

$$A = \frac{x_0(1-x_0)}{B(\alpha,2-\alpha)}\int_{x_0}^{1}dx'\Big[G_\tau(x|x') - G_\tau(x|x_0) - (x'-x_0)\partial_{x_0}G_\tau(x|x_0)\Big](x'-x_0)^{-1-\alpha}(1-x')^{\alpha-1}, \tag{D12}$$

$$B = \frac{x_0(1-x_0)}{B(\alpha,2-\alpha)}\int_{0}^{x_0}dx'\Big[G_\tau(x|x') - G_\tau(x|x_0) - (x'-x_0)\partial_{x_0}G_\tau(x|x_0)\Big](x_0-x')^{-1-\alpha}x'^{\alpha-1}. \tag{D13}$$

Note that the integrals A and B are convergent for $\alpha \in (0,2)$, because, near $x' \sim x_0$, the terms inside $[\cdots]$ are $\mathcal{O}\big((x'-x_0)^2\big)$ and so the integrands are $\mathcal{O}\big(|x'-x_0|^{1-\alpha}\big)$. The jump kernel is given by

$$w(x'|x_0) = \begin{cases} \frac{x_0(1-x_0)}{B(\alpha,2-\alpha)}(x'-x_0)^{-1-\alpha}(1-x')^{\alpha-1} & (x' > x_0) \\ \frac{x_0(1-x_0)}{B(\alpha,2-\alpha)}(x_0-x')^{-1-\alpha}(x')^{\alpha-1} & (x' < x_0). \end{cases} \tag{D14}$$

When $1 < \alpha < 2$, this density agrees with Equation 21 of the main text (up to a proportionality constant). The advection is given by

$$\begin{aligned} V(x_0) &= -\mathrm{PV}\int_0^1 dx' w(x'|x_0)(x'-x_0) \\ &= \frac{x_0(1-x_0)}{B(\alpha,2-\alpha)}\Big(\int_0^{x_0-0}dx'(x_0-x')^{-\alpha}x'^{\alpha-1} - \int_{x_0+0}^{1}dx'(x'-x_0)^{-\alpha}(1-x')^{\alpha-1}\Big). \end{aligned} \tag{D15}$$

Note that, when $\alpha > 1$, the limit $\lim_{\epsilon\to 0}\int_0^{x_0-\epsilon} + \int_{x_0+\epsilon}^1$ in Equation D15 does not exist, although this divergence is rather formal since there exists a natural cutoff $\epsilon \sim \frac{1}{N}$ for a finite-size population.

Appendix E: Analytic results in the marginal case $\alpha = 1$

Although the main target of our present study is the case of $1 < \alpha < 2$, we here provide analytical results for $\alpha = 1$, which have not been derived before.

1. Site frequency spectrum in the presence of genuine selection

The transition density for $\alpha = 1$ in the presence of natural selection is derived in [32] (see [27] for neutral case). In x space, it is given by

$$G^{\sigma}(x, \tau|x_0) = \frac{1}{2\pi x(1-x)} \frac{\sin \pi\eta}{\cos \pi\eta + \cosh[\eta \log \frac{e^{\sigma}x}{1-x} - \log \frac{e^{\sigma}x_0}{1-x_0}]}, \tag{E1}$$

where $\eta \equiv e^{-\tau}$ and σ is the selective advantage (there is an erratum in Equation 38 in [32]).

For the purpose of computing the site frequency spectrum (or, equivalently, the mean sojourn time), we set $x_0 = 1/N$. Since we are considering the large N limit, the denominator of Equation E1 can be rewritten as

$$\begin{aligned} \cos \pi\eta + \cosh[\eta \log \frac{e^{\sigma}x}{1-x} - \log \frac{e^{\sigma}x_0}{1-x_0}] &\approx \frac{1}{2} \exp[\eta \log \frac{x}{1-x} - \sigma(1-\eta) + \log N] \\ &= \frac{1}{2} N (\frac{x}{1-x})^{\eta} e^{-\sigma(1-\eta)}. \end{aligned} \tag{E2}$$

Thus, the transition density for $x_0 = \frac{1}{N}$ can be written as

$$G^{\sigma}(x, \eta|x_0 = \frac{1}{N}) = \frac{e^{\sigma}}{\pi N x(1-x)} \frac{\sin \pi\eta}{(\frac{x}{1-x}e^{\sigma})^{\eta}}. \tag{E3}$$

Near the boundaries, this can be approximated as

$$G^{\sigma}(x, \eta|x_0 = \frac{1}{N}) = \begin{cases} \frac{e^{\sigma}}{\pi N x} \frac{\sin \pi\eta}{(xe^{\sigma})^{\eta}} & (x \ll 1) \\ \frac{e^{\sigma}}{\pi N(1-x)} ((1-x)e^{-\sigma})^{\eta} \sin \pi\eta & (1-x \ll 1). \end{cases} \tag{E4}$$

The site frequency spectrum is given by $f_{\mathrm{SFS}}(x) = N\mu \times t(x)$, where μ is the mutation rate per generation, and $t(x)$ is the mean sojourn time density, which is given by

$$\begin{aligned} t(x) &= \int_0^{\infty} dt\, G^{\sigma}(x, \tau|x_0 = \frac{1}{N}) = \int_0^1 \frac{d\eta}{\eta} G^{\sigma}(x, \eta|x_0) \\ &= \begin{cases} \frac{e^{\sigma}}{\pi N x} \int_0^1 \frac{d\eta}{\eta} \frac{\sin \pi\eta}{(xe^{\sigma})^{\eta}} & (x \ll 1) \\ \frac{e^{\sigma}}{\pi N(1-x)} \int_0^1 \frac{d\eta}{\eta} \sin(\pi\eta) \left((1-x)e^{-\sigma}\right)^{\eta} & (1-x \ll 1). \end{cases} \end{aligned} \tag{E5}$$

Next, we compute the integrals in Equatiion E5, asymptotically close to the absorbing boundaries (see Equation E14 for the final results). To evaluate Equation E5 for $x \ll 1$, we first consider the integral,

$$I_{\epsilon} = \int_0^1 d\eta \exp f(\eta). \tag{E6}$$

When $f(\eta)$ has a sharp peak at $\eta = \eta^*$, we approximate this integral as

$$I_{\epsilon} \approx e^{f(\eta^*)} \sqrt{\frac{2\pi}{|f''(\eta^*)|}}. \tag{E7}$$

In our case,

$$f(\eta) = -\log \eta - \log(\epsilon)\, \eta + \log \sin \pi\eta \tag{E8}$$

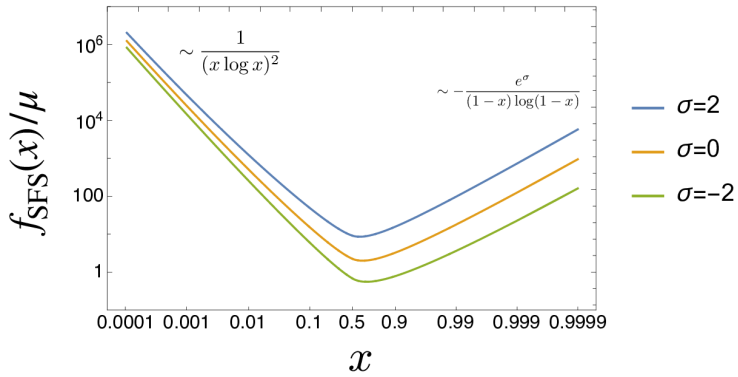

Figure 16. The SFS $f_{\rm SFS}(x)/\mu$ when $\alpha = 1$ for the selective advantage $\sigma = -2, 0, 2$. $f_{\rm SFS}(x)$ is obtained by numerically evaluating the exact expression of $t(x)$ in the first line of Equation E5. As $x \to 0$, $f(x)$ becomes independent of σ. Near $x = 1$, while the magnitude of $f(x)$ depends on σ, the scaling behavior (slope in the log-log plot) does not. See Equation E14.

where $\epsilon = xe^\sigma$. $f(\eta)$ takes the maximum value at $\eta = \eta^* \approx 1 + \frac{1}{\log\epsilon}$ [1]. At $\eta = \eta^*$, $f(\eta^*) \approx -\log\epsilon - 1 + \log\frac{-\pi}{\log\epsilon}$ [2], and $f''(\eta^*) \approx 1 - \frac{\pi^2}{\sin^2(\frac{\pi}{\log\epsilon})} \approx \log^2\epsilon$. The saddle-point evaluation in Equation E7 is precise when $\epsilon \ll 1$. By using these expressions, I_ϵ can be evaluated as

$$I_\epsilon \approx \frac{\sqrt{2\pi}}{-\log\epsilon} e^{-1} \frac{-\pi}{\epsilon\log\epsilon} = \frac{\sqrt{2\pi}\pi e^{-1}}{\epsilon\log^2\epsilon}. \tag{E9}$$

By setting $\epsilon = xe^\sigma$, we find

$$f_{\rm SFS}(x \approx 0) \sim \mu\frac{\sqrt{2\pi}e^{-1}}{(x(\log x + \sigma))^2} \sim \mu\frac{\sqrt{2\pi}e^{-1}}{(x\log x)^2} \propto \mu\frac{1}{(x\log x)^2}. \tag{E10}$$

Next, to evaluate Equation E5 for the high-frequency end, we consider the following integral

$$I'_\epsilon = \int_0^1 \frac{d\eta}{\eta}\sin(\pi\eta)\epsilon^\eta. \tag{E11}$$

When $\epsilon \ll 1$, the integrand takes the maximum value at the boundary $\eta = 0$. Thus,

$$I'_\epsilon \approx \int_0^1 d\eta\ \pi\epsilon^\eta = \frac{\pi(-1+\epsilon)}{\log\epsilon} \approx \frac{-\pi}{\log\epsilon}. \tag{E12}$$

By setting $\epsilon = (1-x)e^{-\sigma}$, we find

$$f_{\rm SFS}(x \approx 1) \sim -\mu\frac{e^\sigma}{(1-x)(\log(1-x) - \sigma)} \sim -\mu\frac{e^\sigma}{(1-x)\log(1-x)}. \tag{E13}$$

In summary, the SFS in Equation E5 is given by

$$f_{\rm SFS}(x) \sim \begin{cases} \mu\frac{1}{(x\log x)^2} & (\text{for } x \ll 1, e^{-\sigma}) \\ -\mu\frac{e^\sigma}{(1-x)\log(1-x)} & (\text{for } 1-x \ll 1, e^\sigma). \end{cases} \tag{E14}$$

Note that the dependence on σ disappears when $x \ll 1$. Figure 6 shows the plots of the SFS.

---

[1] η^* is obtained from $0 = f'(\eta^*) = -\frac{1}{\eta^*} - \log(\epsilon) + \frac{\pi}{\tan\pi\eta^*} \approx -\log(\epsilon) + \frac{\pi}{\tan\pi\eta^*} \approx -\log(\epsilon) + \frac{1}{1-\eta^*}$.

[2] Although the magnitudes of -1 and $\log\frac{-\pi}{\log\epsilon}$ are small compared to $-\log\epsilon$, we need to retain these two terms because $f(\eta^*)$ contributes to I_ϵ through $e^{f(\eta^*)}$.

For comparison, we write the site frequency spectrum for the Wright-Fisher model ($\alpha \geq 2$) (see, for example, [3, 51]);

$$f_{\rm SFS}^{\rm WF}(x) = \theta \frac{e^{2\sigma}(1-e^{-2\sigma(1-x)})}{(e^{2\sigma}-1)x(1-x)}. \tag{E15}$$

The asymptotic forms near the boundaries are given by

$$f_{\rm SFS}^{\rm WF}(x) \approx \begin{cases} \theta \frac{1}{x} \quad (\text{for } x \ll 1) \\ \theta\sigma(1+\coth\sigma)(1+(\sigma-1)(x-1)), \quad (\text{for } 1-x \ll 1, |\sigma|(1-x) \ll 1) \end{cases}$$

where we have expanded the SFS around $x = 1$ up to the sub-leading order. For a sufficiently strong selection ($\sigma > 1$), the SFS increases with x at the high-frequency end. However, unlike the case of $\alpha < 2$, the increase is not strong and the SFS approaches the constant $\sigma(1 + \coth\sigma)$ as $x \to 1$.

2. Dynamics of the median of allele frequencies

When $\alpha = 1$, we can derive a simple differential equation that described the median of trajectories. In the logit space, the transition density is given by

$$G(\psi, \rho|\psi_0) = \frac{\sin \pi\rho}{2\pi\{\cos \pi\eta + \cosh\left[\rho(\psi+\sigma) - (\psi_0+\sigma)\right]\}} \tag{E16}$$

where $\rho = e^{-\tau}$. The median $\Psi^{\rm med}$ (at a given time point ρ) is characterized by

$$\int_{-\infty}^{\Psi^{\rm med}} G(\psi, \rho|\psi_0) d\psi = \frac{1}{2}. \tag{E17}$$

From the symmetry of cosh, the median is given by the peak of the transition density;

$$\Psi^{\rm med} = -\sigma + \frac{1}{\rho}(\psi_0 + \sigma). \tag{E18}$$

By differentiating Equation E18 with respect to ρ and eliminating ψ_0, we obtain

$$\frac{d}{d\rho}\Psi^{\rm med} = -\frac{1}{\rho^2}(\psi_0+\sigma) = -\frac{1}{\rho}(\Psi^{\rm med}+\sigma). \tag{E19}$$

Noting that $\frac{d}{dt} = -\rho\frac{d}{d\rho}$, we find

$$\frac{d}{d\tau}\Psi^{\rm med} = \Psi^{\rm med} + \sigma. \tag{E20}$$

Since the median is invariant under a coordinate transformation, the median $X^{\rm med}$ in the x space is simply related with $\Psi^{\rm med}$ via the logit transformation, $\log \frac{X^{\rm med}}{1-X^{\rm med}} = \Psi^{\rm med}$. By differentiating this with respect to time and using Equation E20, we obtain

$$\frac{d}{d\tau}X^{\rm med} = X^{\rm med}(1-X^{\rm med})(\log\frac{X^{\rm med}}{1-X^{\rm med}} + \sigma). \tag{E21}$$

3. Allele frequency dynamics conditioned on fixation

By using Bayes' theorem, the probability distribution of the allele frequency conditioned on fixation can be written as

$$P(x, \tau|x_0, \text{fixation}) = P(x, \tau, \text{fixation}|x_0) \times \frac{1}{P(\text{fixation}|x_0)} \tag{E22}$$

$$= P(x, \tau|x_0) \times \frac{P(\text{fixation}|x)}{P(\text{fixation}|x_0)}. \tag{E23}$$

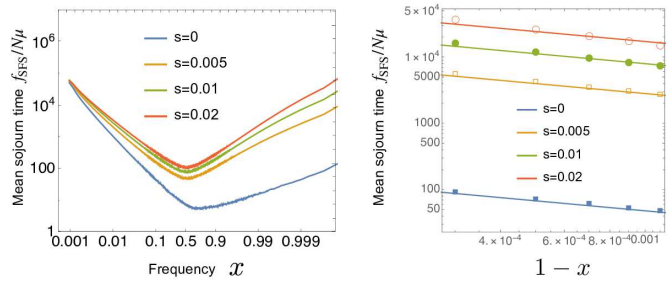

Figure 17. Left: The SFS under positive selection $s = 0,\ 0.005,\ 0.01,\ 0.02$. $\alpha = 1.5$ and $N = 10^6$. Right: The SFS near $x = 1$. The straight lines are drawn assuming $SFS(x) \propto 1/(1-x)^{2-\alpha}$. The slope is almost independent of s.

The fixation probability for the initial frequency x_0 is given by (see [32])

$$P(\text{fixation}|x_0) = \frac{x_0 e^{\sigma}}{1 + x_0(e^{\sigma} - 1)}. \tag{E24}$$

In particular, the fixation probability of a single mutant is given by

$$P(\text{fixation}|x_0 = \frac{1}{N}) \sim \frac{1}{N^{1-s}}. \tag{E25}$$

By using Equation E24, the conditioned probability in Equation E23 is computed as

$$\begin{aligned} P(x, t|x_0, \text{fixation}) &= \frac{1}{2\pi x(1-x)} \times \frac{x e^{\sigma}}{1 + x(e^{\sigma} - 1)} \times \frac{1 + x_0(e^{\sigma} - 1)}{x_0 e^{\sigma}} \\ &\times \frac{\sin \pi\rho}{\cos \pi\rho + \cosh[\rho \log \frac{e^{\sigma} x}{1-x} - \log \frac{e^{\sigma} x_0}{1-x_0}]} \\ &= \frac{1}{2\pi x_0(1-x)} \frac{1 + x_0(e^{\sigma} - 1)}{1 + x(e^{\sigma} - 1)} \frac{\sin \pi\rho}{\cos \pi\rho + \cosh[\rho \log \frac{e^{\sigma} x}{1-x} - \log \frac{e^{\sigma} x_0}{1-x_0}]}. \end{aligned} \tag{E26}$$

Appendix F: Site frequency spectra in presence of selection

Here, we argue the effect of the genuine selection on the SFS by using the effective bias when $1 < \alpha < 2$. As discussed in the main text, there is a crossover point x_c, shown in Equation 41, below which the selection is negligible compared to the effective bias (see Figure 13). Thus, we can expect that the SFS becomes independent of the selective advantage σ for a sufficiently small frequency x. Similarly, for the high-frequency end $1 - x \ll 1$, the selection is negligible compared with the effective bias. Therefore, we expect that $f_{\text{SFS}}(x) \sim \frac{1}{V_{\text{eff}}(x)} \propto (1-x)^{-\alpha+2}$ even in the presence of natural selection. In particular, the exponent is independent of σ. Figure 17 shows the numerical results when $\alpha = 1.5$. As x approaches 0, the SFS becomes independent of the selective advantage σ. For frequent variants $1 - x \ll 1$, the SFS can be fitted well by $(1-x)^{-\alpha+2}$, while the magnitude of the SFS increases with σ. A similar result can be obtained analytically when $\alpha = 1$ (see Appendix E).

Appendix G: Derivation of the rate of adaptation in Equation 48 of the main text

Here, we conjecture the rate of adaptation for an asexual population with a broad offspring distribution ($1 < \alpha < 2$) in the clonal-interference regime, using a self-consistency condition argument described in [48].

We assume that mutations have a fixed effect s much larger than the mutation rate μ_{B} at which they arise. First, we consider the dynamics of the fittest sub-population that becomes established at the nose of the fitness wave. We can estimate the size of the sub-population when established from the establishment probability of a single fittest mutant;

$$N_{est} \sim \frac{1}{P_{\text{fix}}(qs)}, \tag{G1}$$

where qs ($q \in \mathbb{N}$) is the fitness lead of the sub-population compared with the mean of the whole population, and the fixation probability is given by Equation 42, $P_{\rm fix} \sim (qs)^{\frac{1}{\alpha-1}}$. In the time this sub-population is seeded and becomes established, the mean fitness should increase by s. This implies that, after its establishment, this sub-population will initially grow exponentially at rate $(q-1)s$. The growth rate will slow down to 0 when it fixes. Therefore, the time from establishment to fixation can be estimated as

$$t_{\rm fix} \sim \frac{1}{(q-1)s/2} \ln \frac{N}{N_{est}} = \frac{1}{(q-1)s/2} \ln N P_{\rm fix}(qs) \tag{G2}$$

where $(q-1)s/2$ is its average growth rate between the establishment and fixation. Thus, the rate of adaptation is given by

$$R = \frac{(q-1)s}{t_{\rm fix}} \sim \frac{((q-1)s)^2}{2 \ln N P_{\rm fix}(qs)}. \tag{G3}$$

Second, we focus on successive events of establishments at the edge of the fitness wave. We define t_{est} as the mean time interval between two successive establishments. An established sub-population grows like $n(t) \sim N_{est} e^{(q-1)st}$, from which the next event of establishment is produced with rate $n(t)\mu_{\rm B} P_{\rm fix}(qs)$. Therefore, t_{est} can be estimated from

$$\mu_{\rm B} P_{\rm fix}(qs) \int_0^{t_{est}} n(t) dt \approx 1, \tag{G4}$$

which leads to $t_{est} \sim \frac{1}{(q-1)s} \ln[\frac{s}{\mu_{\rm B}}]$. Since the nose of the fitness wave advances at a speed $R = \frac{s}{t_{est}}$, we have

$$R = \frac{s}{t_{est}} \sim \frac{(q-1)s^2}{\ln \frac{s}{\mu_{\rm B}}}. \tag{G5}$$

By comparing Equations G3 and G5, we obtain

$$q \sim 1 + \frac{2 \ln(N P_{\rm fix}(qs))}{\ln \frac{s}{\mu_{\rm B}}}, \quad R \sim \frac{2s^2 \ln(N P_{\rm fix}(qs))}{(\ln \frac{s}{\mu_{\rm B}})^2}. \tag{G6}$$

By substituting $P_{\rm fix} \sim (qs)^{\frac{1}{\alpha-1}}$ into Equation G6, we obtain

$$q \sim 1 + \frac{2 \ln(N s^{\frac{1}{\alpha-1}})}{\ln \frac{s}{\mu_{\rm B}}}, \; R \sim \frac{2s^2 \ln(N s^{\frac{1}{\alpha-1}})}{(\ln \frac{s}{\mu_{\rm B}})^2}, \tag{G7}$$

where we used $\ln N q^{\frac{1}{\alpha-1}} \approx \ln N$. In the limit $\alpha \to 2$, the above results reproduce those in [48].

The case of $\alpha = 1$ can be discussed in a similar way. Suppose that the population is monoclonal. The fixation probability of a mutant is given by $P_{\rm fix} \sim N^{-1+s}$ (see Equation E25), which implies that the establishment size is roughly given by $N_{est} \sim N^{1-s}$. While the timescale of establishment of a mutant is given by $(\mu_{\rm B} N P_{\rm fix})^{-1} = (\mu_{\rm B} N^s)^{-1}$, the timescale of fixation is given by $t_{\rm fix} \sim \frac{1}{s} \log \frac{N}{N_{est}} \sim \log N$. Thus, the successive selection sweeps occur if $(\mu_{\rm B} N^s)^{-1} \gg \log N$, or equivalently,

$$\mu_{\rm B} N^s \log N \ll 1 \quad \text{(successive selective sweeps)}. \tag{G8}$$

By substituting $P_{\rm fix} \sim N^{-1+s}$ into Equation G6, the rate of adaptation in the clonal-interference regime is given by

$$R \sim \frac{2s^3 \ln N}{(\ln \frac{s}{\mu_{\rm B}})^2}. \tag{G9}$$

In the successive-sweeps regime, the adaptation rate is given by

$$R = s\mu_{\rm B} N \times P_{\rm fix}(s) \sim s\mu_{\rm B} N^s. \tag{G10}$$

Note that clonal interference becomes unlikely to occur as the offspring distribution becomes broader. For example, when $\alpha = 1$, the population size needs to be $N \gg 10^{41}$ for $\mu_{\rm B} = 10^{-4}, s = 0.05$ to satisfy $\mu_{\rm B} N^s \log N \gg 1$.

Figure 18 shows the numerical results of the adaptation rate R versus the selection coefficient s. The parameters used in the simulation are in the regime of clonal interference. When $1 < \alpha$, R is approximately proportional to s^2, while, when $\alpha = 1$, R is approximately proportional to s^3, which are consistent with Equations G6 and G9. However, when $\alpha = 1$, the quantitative agreement between the numerical result and the theoretical prediction is not good, and a further investigation is needed to validate Equation G9.

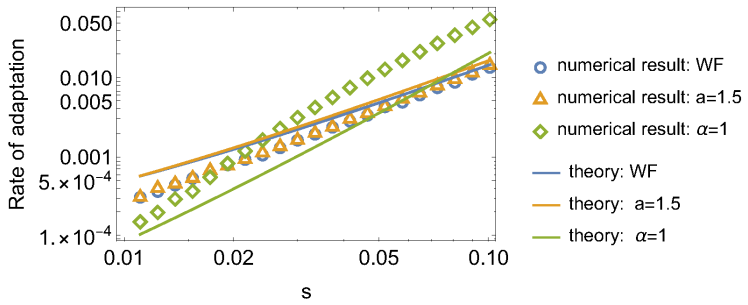

Figure 18. The open markers show the numerical results of R as a function of s, while the curves show the theoretical predictions, based on the heuristic argument. The m rate of beneficial mutations is $\mu = 10^{-4}$. The population size is $N = 10^{100}$ for $\alpha = 1$,$N = 10^{10}$ for $\alpha = 1.5$, and $N = 10^{8}$ for the Wright-Fisher model.

Appendix H: Stationary distributions of traveling wave model in the presence of natural selection

In Figure 14 of the main text, the mutant allele is assumed be neutral. Here, we provide the results in the case where mutants have a fitness advantage σ (Figure 19). As in the main text, symmetrically reversible mutations are assumed.

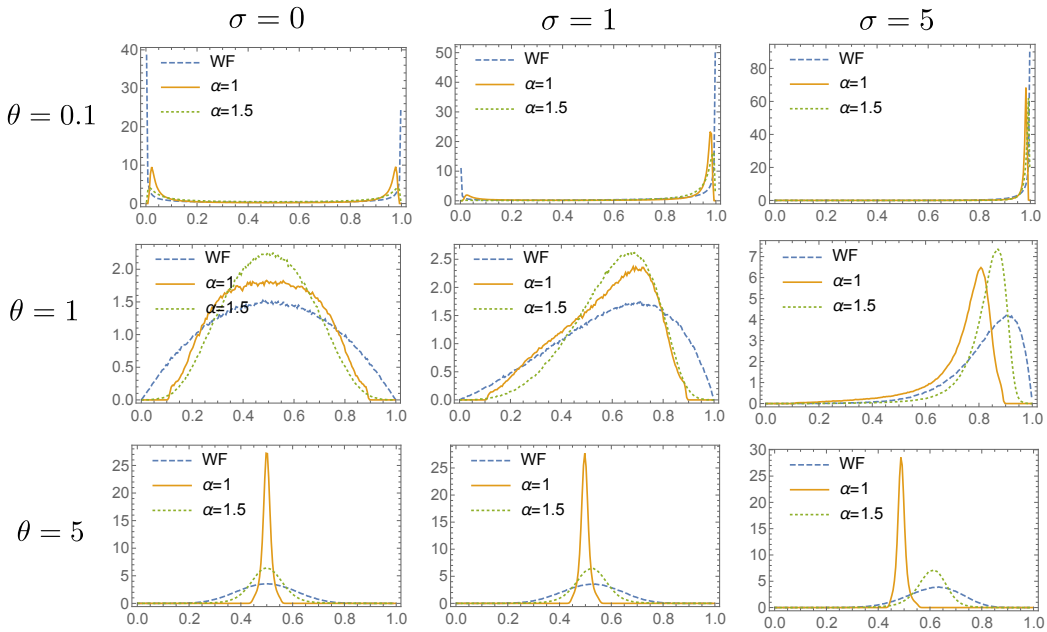

Figure 19. The stationary distributions of the mutant frequency for $\theta = 0.1, 1, 5$. $\sigma = 0, 1, 5$. σ is the selection coefficient in the time-continuous description, $\sigma = sT_c$.

Appendix I: Numerical simulations

Simulations are implemented in C++ with the GNU scientific library's random number generators. Results obtained from the simulations are analyzed by Mathematica. The codes are freely available upon request.

1. Numerical synthesis of Pareto random variables and α-stable distribution

In order to generate the mutant frequency of the gamete pool, we need to compute the sums of random Pareto variables,

$$M = \sum_{i=1}^{Nx} u_i, \quad W = \sum_{i=1}^{N(1-x)} v_i, \tag{I1}$$

where u_i, v_i are drawn from the Pareto distribution $P_U(u) = \alpha/u^{\alpha+1}$ $(u \geq 1)$. One simple way to synthesize u_i, v_i is to sample a number r from the uniform distribution on $(0, 1)$ and compute $r^{-\frac{1}{\alpha}}$.

To generate the sums M, N efficiently for large N (e.g. $N \sim 10^6$), we can use the generalized central limit theorem when xN and $(1-x)N$ are large. In simulations, when $xN < 100$, M is generated directly by synthesizing xN random variables $\{u_i\}$, while, when $xN \geq 100$, M is generated by sampling a random number ζ from the α-stable distribution and then determining $M = \sum_i u_i$ from Equation A1. W is generated in a similar way.

After generating M and W, the population is updated by the binomial sampling with the success probability $p = \frac{M}{M+W}$ (although this sampling process can be omitted when $\alpha \geq 2$ since the fluctuations associated with the binomial sampling is negligible compared to the fluctuations associated with M and N). Natural selection and mutations are implemented by modifying the success probability $p = \frac{M}{M+W}$ as

$$\frac{p(1+s)}{p(1+s)+(1-p)}(1-\mu_{M\to W}) + \frac{(1-p)}{p(1+s)+(1-p)}\mu_{W\to M}, \tag{I2}$$

where $\mu_{W\to M}$ is the mutation rate from the wild-type to the mutant allele, and $\mu_{M\to W}$ is the mutation rate in the reverse direction.

2. Site frequency spectrum

Since the SFS is proportional to the mean sojourn time, the SFS can be computed numerically by generating trajectories staring with $x_0 = \frac{1}{N}$ until fixation or extinction and measuring how many times a trajectory visits a given frequency interval on average.

3. Numerical simulation of the model of range expansion in the main text

We first review the numerical implementation of the range expansion model with two neutral alleles without mutations [21]. The per capita growth rate $r(n)$ with an Allee effect is given by

$$r(n) = r_0(1 - \frac{n}{K})(1 + B\frac{n}{K}), \tag{I3}$$

where $n = n_1 + n_2$ is the sum of the two population densities, and B is the strength of cooperativity. In each deme, there are three types; allele 1, allele 2, and "empty". At each time step, the configuration of deme x is updated by the trinomial sampling process with

$$p_i = \frac{\tilde{n}_i}{K(1 - r(\tilde{n})\tau)} \text{ for } i = 1, 2 \text{ and } p_{\text{empty}} = 1 - p_1 - p_2, \tag{I4}$$

where $\tilde{n}_i$ is the population density after migration,

$$\tilde{n}_i(t, x) = \frac{m}{2}n_i(t, x-a) + (1-m)n_i(t, x) + \frac{m}{2}n_i(t, x+a), \tag{I5}$$

and $\tilde{n}$ in the denominator of Equation I4 is the sum of these densities, $\tilde{n} = \tilde{n}_1 + \tilde{n}_2$, and a denotes the width of a deme. The expectation value of the total density n after one time step is given by

$$K \sum_{i=1,2} p_i = \frac{\tilde{n}}{1 - r(\tilde{n})\tau} \approx \tilde{n}(1 + r(\tilde{n})\tau), \tag{I6}$$

which explains the denominator of Equation I4. In the simulation, $a = 1$ and $\tau = 1$ are used.

As in the standard Wright-Fisher model, a mutation process can be introduced by using the success probabilities $\mathbf{p}' = (p_1', p_2')^T$ given by

$$\mathbf{p}' = U\mathbf{p}, \tag{I7}$$

where $\mathbf{p} = (p_1, p_2)^T$ and U is a matrix representing mutational transitions. In the case of symmetrical mutations in the main text, U is given by

$$U = \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix}, \tag{I8}$$

This model serves as a microscopic description of our (non-spatial) macroscopic model of the population with a broad offspring distribution $p(U = u) \sim \frac{1}{u^{\alpha+1}}$. We can argue the relation between the parameters in the two models by comparing the coalescent timescales. As established in [21], for a semi-pushed wave ($2 < B < 4$), the coalescent timescale is given by

$$T_c^{\text{micro}} \sim N^{2\frac{\sqrt{1-\gamma(B)^2}}{1-\sqrt{1-\gamma(B)^2}}}. \tag{I9}$$

where $\gamma(B) = \frac{v_F}{v} = 2(\sqrt{\frac{B}{2}} + \sqrt{\frac{2}{B}})^{-1}$ is the ratio of the Fisher velocity $v_F = 2\sqrt{Dr_0}$ to the wave velocity $v = \sqrt{r_0 D}(\sqrt{\frac{B}{2}} + \sqrt{\frac{2}{B}})$. On the other hand, the coalescent timescale T_c^{macro} in the macroscopic description for $1 < \alpha < 2$ is proportional to $N^{\alpha-1}$ (see Equation 15). By comparing the exponents, a semi-pushed wave with B corresponds to the macroscopic model with [3]

$$\alpha = 2\frac{\sqrt{1-\gamma(B)^2}}{1-\sqrt{1-\gamma(B)^2}} + 1. \tag{I10}$$

For example, $B = 3$ corresponds to $\alpha = 1.5$. In addition, the mutation rate μ_{micro} per generation in the microscopic model and the mutation rate μ_{macro} per generation in the macroscopic model should be related by $\mu_{\text{micro}} \times T_c^{\text{micro}} \sim \mu_{\text{macro}} \times T_c^{\text{macro}}$.

In the three panels (Left. Center, Right) in Figure 14B of the main text, The following parameters are used.

- Left: $B = 1,\ \mu = (5\times10^{-4},\ 5\times10^{-5}),\ K = 28000$ for the microscopic model model, and $\alpha = 1,\ \theta = (1.5,\ 0.15)$ for the macroscopic model.
- Center: $B = 3,\ \mu = (2\times10^{-4},\ 2\times10^{-5}),\ K = 35000$ for the microscopic model, and $\alpha = 1.5,\ \theta = (1.6,\ 0.16)$ for the macroscopic model.
- Right: $B = 8,\ \mu = (1\times10^{-5},\ 1\times10^{-6}),\ K = 57000$ for the microscopic model, and the Wright-Fisher model, $\theta = (2.4,\ 0.24)$ for the macroscopic model.

In all of the three cases, the growth rate $\frac{r_0}{\tau} = 0.01$ and the migration probability $m = \frac{2D\tau}{a^2} = 0.125$ are used in the microscopic model, and the population size $N = 10^5$ is used in the macroscopic model. Note that, to compare the microscopic model with the macroscopic model, the value of the carrying capacity K for each case is chosen such that the size of the front population $\frac{K}{k}$, where k is the spatial decay rate of the population density[4], approximately agrees with the population size $N = 10^5$ in the macroscopic model.

Appendix J: Areas swept by trajectories

1. A scaling argument on area distributions

Consider frequency trajectories that depart from a single mutant $x_0 = \frac{1}{N}$ and are eventually absorbed either at $x = 0$ or at $x = 1$. For each of such trajectories, we can define the area in $x - \tau$-space swept by the trajectory (see

[3] Note that the definition of the parameter α_H in [21] is different from our definition of α. For $1 < \alpha < 2$, which corresponds to the semi-pushed wave region $-1 < \alpha_H < 0$, the two definitions are related by $-\alpha_H = \alpha - 1$.

[4] $k = \sqrt{\frac{r_0}{D}}$ for $0 < B < 2$, and $k = \sqrt{\frac{r_0 B}{2D}}$ for $B \geq 2$ [21].

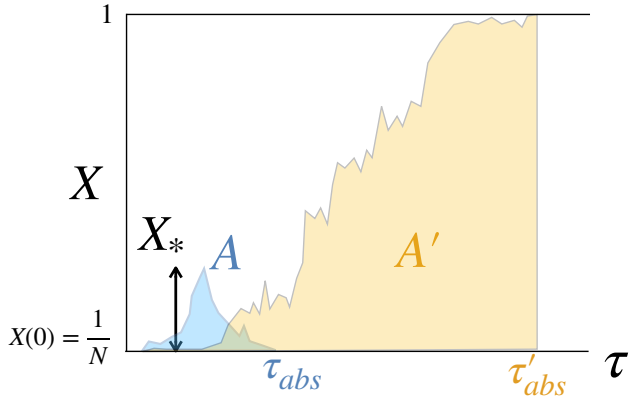

Figure 20. An area A swept by a trajectory that eventually goes extinct and an area A' swept by a trajectories that eventually gets fixed are illustrated. $\tau_{\rm abs}$ and $\tau'_{\rm abs}$ are the extinction time and the fixation time, respectively.

Figure 21. Left: The area distribution $p(A)$ for $\alpha = 1$, 1.5 and the Wright-Fisher model. The straight lines show the scaling-argument predictions, $p(A) \propto \frac{1}{A^{1+\frac{1}{\alpha}}}$. $N = 10^6$. Right: The tail of $p(A)$ in the large-A region.

Figure 20),

$$A = \int_0^{\tau_{\rm abs}} x(\tau) d\tau, \tag{J1}$$

where $\tau_{\rm abs}$ is the absorption time of the trajectory. While this quantity is defined for a population without spatial structure, we expect that it has a natural interpretation in a model of range expansion as a spatial integration over the mutant frequency (i.e., the abundance of the mutant type), since τ in Equation J1 is related with the spatial position of the traveling wave in the comoving frame.

Here, we examine how the area A defined in Equation J1 depends on the exponent α of the offspring distribution. The left panel of Figure 21 shows the numerical results of the area distribution $p(A)$ for $\alpha = 1,\ 1.5$, and the Wright-Fisher model (corresponding to $\alpha \geq 2$). In a wide range of A, areas are distributed according to $p(A) \sim \frac{1}{NA^{1+\frac{1}{\alpha}}}$.

Focusing on small areas, which correspond to extinct trajectories, this power-law behavior can be rationalized again from a scaling argument: First, by using Equation 3, a trajectory whose maximum frequency is $x_* \ll 1$ sweeps an area roughly given by $A \sim x_* \times \tau_{\rm ext} \sim x_*^{\alpha}$ (see Figure 20), i.e., $x_* \sim A^{\frac{1}{\alpha}}$. Second, from the neutrality, the cumulative probability $\Pr(X_* > x_*)$ that a single mutant achieves a frequency larger than x_* before absorption is estimated as $\Pr(X_* > x_*) \sim \frac{1}{Nx_*}$. Hence, the density $p(x_*)$ is given by $p(x_*) = -\frac{d}{dx_*}\Pr(X_* > x_*) \sim \frac{1}{Nx_*^2}$. Combining these two results, we can estimate the area distribution $p(A)$ as

$$p(A) \sim p(x_*)\frac{dx_*}{dA}\bigg|_{x^*=A^{\frac{1}{\alpha}}} \sim \frac{1}{Nx_*^2}(x_*)^{-\alpha+1}\bigg|_{x^*=A^{\frac{1}{\alpha}}} \sim \frac{1}{N}A^{-1-\frac{1}{\alpha}}. \tag{J2}$$

When $\alpha \to 2-0$ (Wright-Fisher limit), the distribution becomes $p(A) \sim \frac{1}{N}A^{-\frac{3}{2}}$, which can be analytically confirmed by solving a backward diffusion equation of the Wright-Fisher diffusion (see Appendix J-2).

The numerical results indicate that, when $1 \leq \alpha < 2$, there is an uptick in the area distribution $p(A)$, which comes from fixed trajectories (see the case of $\alpha = 1$ in the right panel of Figure 21). The uptick becomes less pronounced as α increases. For the Wright-Fisher model, we can analytically prove that $p(A)$ monotonically decreases with A.

2. Area distribution in the Wright-Fisher model

Here, we derive an analytical result of Equation J1 for the Wright-Fisher diffusion process.

Consider a Langevin equation

$$\frac{dX}{d\tau} = v(X) + \xi(\tau), \tag{J3}$$

with $\langle \xi(\tau)\xi(\tau')\rangle = 2D(x)\delta(\tau - \tau')$. Assume the initial value $X(\tau = 0) = x_0 \in (0, 1)$ and the absorbing boundaries at $X = 0, 1$. For a given trajectory departing from x_0 and ending at either one of the boundaries, we consider the "area" defined by

$$A = \int_0^{\tau_{\rm abs}} X(\tau)d\tau \ . \tag{J4}$$

where $\tau_{\rm abs}$ is the absorption time.

The area distribution $\Pi(A; x_0)$ for a given initial condition $X(0) = x_0$ obeys a backward equation. To show this, we discretize the dynamics;

$$\Delta X = vh + W \tag{J5}$$

where h denotes a short time interval and $\langle W_i W_j\rangle = 2Dh\delta_{i,j}$. The transition density is given by

$$T(x_0 + \Delta x|x_0) = \frac{1}{\sqrt{\pi(2D(x_0)h)}} \exp(-\frac{(\Delta x - v(x_0)h)^2}{2(2D(x_0)h)}). \tag{J6}$$

Note that

$$\begin{aligned} \langle \Delta x\rangle_{x_0} &= v(x_0)h, \\ \langle (\Delta x)^2\rangle_{x_0} &= v(x_0)h + 2D(x_0)h. \end{aligned} \tag{J7}$$

By separating a trajectory into the initial step and the remaining part, we have

$$\Pi(A; x_0) = \int d(\Delta x)\ T(x_0 + \Delta x|x_0)\Pi(A - x_0h; x_0 + \Delta x) + o(h), \tag{J8}$$

By Taylor-expanding $\Pi(A - x_0h; x_0 + \Delta x)$, we have

$$\begin{aligned} \Pi(A - x_0h; x_0 + \Delta x) &= \Pi(A; x_0) - \frac{\partial \Pi}{\partial A}x_0h + \frac{\partial \Pi}{\partial x_0}\Delta x \\ &\quad + \frac{1}{2}\frac{\partial^2 \Pi}{\partial A^2}x_0h^2 - \frac{\partial^2 \Pi}{\partial A\partial x_0}x_0h\Delta x + \frac{1}{2}\frac{\partial \Pi}{\partial x_0}\Delta x^2 + \cdots \\ &= \Pi(A; x_0) - \frac{\partial \Pi}{\partial A}x_0h + \frac{\partial \Pi}{\partial x_0}\Delta x + \frac{1}{2}\frac{\partial \Pi}{\partial x_0}\Delta x^2 + o(h), \end{aligned} \tag{J9}$$

Therefore, Equation J8 becomes

$$\Pi(A; x_0) = \Pi(A; x_0) - \frac{\partial \Pi}{\partial A}x_0h + \frac{\partial \Pi}{\partial x_0}\langle \Delta x\rangle_{x_0} + \frac{1}{2}\frac{\partial^2 \Pi}{\partial x_0^2}\langle \Delta x^2\rangle_{x_0} + o(h). \tag{J10}$$

By using Equation J7, we obtain

$$x_0\frac{\partial \Pi}{\partial A} = v(x_0)\frac{\partial \Pi}{\partial x_0} + D(x_0)\frac{\partial^2 \Pi}{\partial x_0^2}. \tag{J11}$$

Although, in the following, we consider the area defined by Equation J4, it can be shown that, for the following integral,

$$\tilde{A} = \int_0^{T^*} dt\ f(X), \tag{J12}$$

the distribution $\Pi(\tilde{A}; x_0)$ satisfies

$$f(x_0)\frac{\partial \Pi}{\partial \tilde{A}} = v(x_0)\frac{\partial \Pi}{\partial x_0} + D(x_0)\frac{\partial^2 \Pi}{\partial x_0^2}. \tag{J13}$$

In the neutral Wright-Fisher model, $v(x_0) = 0$ and $D(x_0) = x_0(1 - x_0)$. The backward equation in Equation J11 is given by

$$\frac{\partial \Pi}{\partial A} = (1 - x_0)\frac{\partial^2 \Pi}{\partial x_0^2}. \tag{J14}$$

From this equation, it follows that $\Pi(A|x_0)$ monotonically decreases with A_0 because the spectrum of the operator $\partial_{x_0}^2 \sim (ik)^2$ is non-positive.

We can determine the area distribution $p(A)$ analytically at least for small A. We are interested in the invasion by a single mutant, $x_0 = \frac{1}{N} \ll 1$. Furthermore, for the purpose of determining the behavior for small areas, we expect that we can ignore the presence of the high-frequency boundary $x = 1$ and solve the problem on the semi-infinite line $x_0 \in (0, \infty)$. Therefore, we consider the following problem:

$$\begin{aligned}
&\frac{\partial \Pi}{\partial A} = \frac{\partial^2 \Pi}{\partial x_0^2}, \\
&\Pi(A; x_0 = 0) = g(A) \\
&\Pi(A = 0; x_0) = 0 \quad \text{for } x_0 > 0 \\
&\lim_{x_0 \to \infty} \Pi(A, x_0) = 0
\end{aligned} \tag{J15}$$

In our case, $g(A) = \delta(A)$, because the trajectory starting from $x_0 = 0$ has $A = 0$.

For a function $f(A)$ of A, we write the Laplace transformation as

$$\hat{f}(s) = \mathcal{L}[f(A)] = \int_0^\infty ds f(t) e^{-sA}. \tag{J16}$$

By take the Laplace transform with respect to A, we have

$$s\hat{\Pi}(s; x_0) = \frac{\partial^2 \hat{\Pi}(s, x_0)}{\partial x_0^2}, \quad \hat{\Pi}(s, 0) = \hat{g}(s). \tag{J17}$$

The solution is

$$\hat{\Pi}(s; x) = e^{-x_0\sqrt{s}}\hat{g}(s). \tag{J18}$$

We take the inverse of the Laplace transformation,

$$\Pi(A; x_0) = \mathcal{L}^{-1}(e^{-x_0\sqrt{s}}\hat{g}(s)). \tag{J19}$$

From the convolution theorem, this is given by the convolution of $\mathcal{L}^{-1}(e^{-x_0\sqrt{s}}) = \frac{x_0}{2\sqrt{\pi}A^{\frac{3}{2}}}e^{-\frac{x_0^2}{4A}}$ and $g(A)$;

$$\Pi(A; x_0) = \int_0^A dA' \frac{x_0}{2\sqrt{\pi}A'^{\frac{3}{2}}} e^{-\frac{x_0^2}{4A'}} g(A - A'). \tag{J20}$$

When $g(A) = \delta(A)$, we have

$$\Pi(A; x_0) = \frac{x_0}{2\sqrt{\pi}A^{\frac{3}{2}}} e^{-\frac{x_0^2}{4A}}. \tag{J21}$$

Especially, when $x_0 = 1/N$, we have

$$\begin{aligned}
\Pi(A; x_0 = \frac{1}{N}) &= \frac{1}{2\sqrt{\pi}NA^{\frac{3}{2}}} e^{-\frac{1}{4AN^2}} \\
&\approx \frac{1}{2\sqrt{\pi}NA^{\frac{3}{2}}},
\end{aligned} \tag{J22}$$

where we have used $e^{-\frac{1}{4AN^2}} \approx 1$ since only areas larger than $x_0 \times d\tau \sim \frac{1}{N} \times \frac{1}{N} = \frac{1}{N^2}$ are meaningful for a finite-size population.

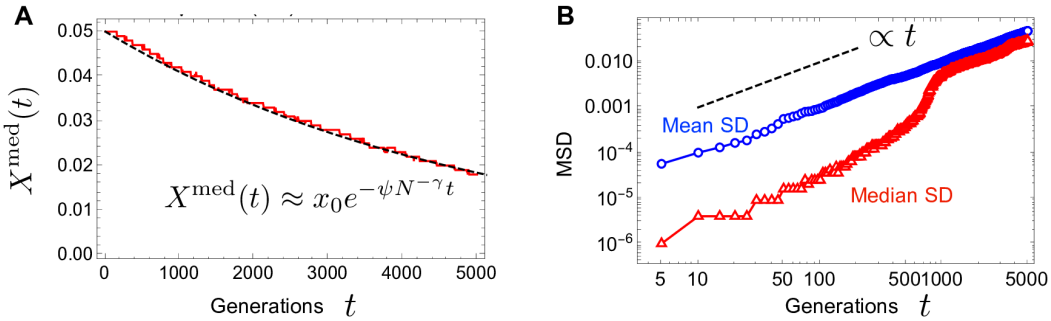

Figure 22. Simulation results of the Eldon-Wakeley model. (A) The median frequency of the Eldon-Wakeley model (red solid) and $X^{\rm med}(t) = x_0 e^{-\psi N^{-\gamma} t}$ (black dashed). $N = 10^3$, $\gamma = 1$, $\psi = 0.1$, $x_0 = 0.05$. (B) The mean and median square displacements (blue and red curves, receptively). The black dashed line $\propto 1/t$ indicates the expectation from the Wright-Fisher (or Moran) model. $N = 10^3$, $\gamma = 1$, $\psi = 0.2$, $x_0 = 0.5$.

Appendix K: Forward-in-time behaviors of the Eldon-Wakeley model

Here, we present simulation results of the median allele frequency and the median and mean square displacements in the Eldon-Wakeley model [11] (see also [52]). As shown below, unlike our model, these quantities do no exhibit sustained power-law behaviors, because of the existence of a characteristic size ψ in the offspring distribution.

We consider the neutral Eldon-Wakeley model, where the following offspring distribution $P_U(u)$ is given by (see Equation (7) in [11]);

$$P_U(u) = (1 - N^{-\gamma})\delta_{u,2} + N^{-\gamma}\delta_{u,\psi N}, \tag{K1}$$

where $\delta_{a,b}$ is the Kronecker delta. $\psi \in (0, 1)$ and the parameters characterizing how large and frequent 'sweepstakes' are.

The limiting process as $N \to \infty$ depends on γ (see Equation (9) in [52]). For $\gamma > 2$, the process is the same as the Wright-Fisher diffusion, while, for $\gamma < 2$, it is described by a jump process whose backward-time generator $\mathcal{L}^\dagger$ is given by

$$L^\dagger P(x, \tau) = xP(x + \psi(1 - x), \tau) - P(x, \tau) + (1 - x)P(x - \psi x, \tau), \tag{K2}$$

where the continuous time τ is related with generations t by $\tau = t/N^\gamma$. The first term of the generator represents a frequency-increasing jump $x \to x + \psi(1 - x)$ with rate x, while the last one represents a frequency-decreasing jump $x \to x - \psi x$ with rate $1 - x$.

Figure 22 shows numerical simulation results for the median of allele frequencies and the median/mean square displacements. The median frequency for a small initial frequency $x_0 \ll 1$ is well described by $X^{\rm med}(t) = x_0 e^{-\psi N^{-\gamma} t}$ (Figure 22A). This exponential decay can be expected from the generator in Equation K2; for $x \ll 1$, frequency-increasing jumps (with rate x) are unlikely to occur, and an allele frequency typically decreases by $-\psi x$ with rate $1 - x \approx 1$. Thus, the median frequency in the Eldon-Wakeley model does not exhibit a power-law behavior.

As for frequency fluctuations, while the mean SD exhibits a normal diffusion as in the Moran (or Wright-Fisher) model, i.e., Mean SD $\propto t$, the median SD does not exhibit a sustained power-law behavior (Figure 22B); in a short- and long time scales, the median SD exhibits a normal diffusion (Median SD $\propto t$), but, for an intermediate timescale ($t \sim 500 - 1000$ generations in the figure), it increases more rapidly than expected from a normal diffusion.