Genealogies of rapidly adapting populations

Richard A. Neher[a,1] and Oskar Hallatschek[b]

[a]Evolutionary Dynamics and Biophysics Group, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; and [b]Biophysics and Evolutionary Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany

The genetic diversity of a species is shaped by its recent evolutionary history and can be used to infer demographic events or selective sweeps. Most inference methods are based on the null hypothesis that natural selection is a weak or infrequent evolutionary force. However, many species, particularly pathogens, are under continuous pressure to adapt in response to changing environments. A statistical framework for inference from diversity data of such populations is currently lacking. Towards this goal, we explore the properties of genealogies in a model of continual adaptation in asexual populations. We show that lineages trace back to a small pool of highly fit ancestors, in which almost simultaneous coalescence of more than two lineages frequently occurs. Whereas such multiple mergers are unlikely under the neutral coalescent, they create a unique genetic footprint in adapting populations. The site frequency spectrum of derived neutral alleles, for example, is nonmonotonic and has a peak at high frequencies, whereas Tajima's D becomes more and more negative with increasing sample size. Because multiple merger coalescents emerge in many models of rapid adaptation, we argue that they should be considered as a null model for adapting populations.

coalescent theory | demographic inference | pathogen evolution | population genetics

Evolutionary change is usually too slow to be observed in real time. A sequence sample represents a static snapshot from which we want to learn about a dynamic evolutionary process. The predominant framework to analyze such population genetic data and infer demographic history is Kingman's neutral coalescent. Within this model, all individuals are equivalent (i.e., there are no fitness differences), and pairs of lineages merge at random. The statistical properties of genealogies in this simple population genetic model can be computed exactly (1, 2), facilitating comparison with data. One central prediction of the neutral coalescent is that the genetic diversity of a population is proportional to its size. This prediction, however, is at odds with the observed weak correlation between genetic diversity and population size, a paradox often remedied by the definition of an effective population size proportional to the genetic diversity. The model has been generalized to account for historic changes in population size, mutation rates, geographical structure, and effects of purifying selection (3–7). Positive selection, however, has proved difficult to incorporate, and progress has been limited to rare selective sweeps (8, 9) and weak selection (10).

In many populations, particularly large microbial populations, selection is neither rare nor weak. Instead, these populations are under sustained pressure to adapt to changing environments. Prominent examples include pathogens like influenza that continuously evade human immune responses or HIV, which establishes a chronic infection despite heavy immune predation. The genealogical trees reconstructed from sequence samples often suggest substantial departure from neutrality; ref. 11 has examples from viral evolution, and ref. 12 has eukaryotic examples. The influenza tree shown in Fig. 1, for instance, is incompatible with a neutral genealogy, because there are parts where many lineages merge in a very brief period, and the tree often branches extremely unevenly, with very few individuals on one branch and many individuals on the other branch. These two observations represent fundamental deviations from the standard neutral model, even when a varying population size is allowed. Strelkova and Lässig

(14) present a detailed analysis of Influenza A evolution and conclude that influenza is governed by coalescence processes different from the Kingman's coalescent.

To analyze and interpret genealogies of populations under sustained directional selection, an alternative simple null model would be extremely useful. The features of genealogies discussed above are, in fact, common to a class of non-Kingman's coalescence models, which have received considerable attention in the mathematical coalescent literature (15, 16). A special case is the Bolthausen–Sznitman coalescent (BSC) (17), which has been shown to describe the genealogies in models where a population expands into uninhabited territory (18). On the basis of a particular exact solution and a phenomenological theory, Brunet et al. (18) conjectured that genealogies in all models of the same universality class [the class of stochastic Fisher–Kolmogorov–Petrovskii–Piscounov (FKPP) waves] (19, 20) are described by the BSC (recent review in ref. 21). This universality class contains all models with short-range dispersal and logistic growth with constant rate in partially filled demes.

We will argue in this article that the BSC emerges generically in models of rapidly adapting asexual populations in a similar way as it describes genealogies in traveling waves of FKPP type. We present extensive computer simulations and investigate the distribution of heterozygosity in the population, the average time to the most recent common ancestor, and the site frequency spectrum (SFS). Most notably, the SFS is nonmonotonic with a large number of high frequency-derived alleles. We then study a simplified model analytically and show that the underlying genealogical process is approximately the BSC. In the discussion, we outline the basic features of the BSC and discuss its applicability to wider classes of models.

Model

The evolutionary dynamics of a large population are mainly determined by the distribution of fitness in the population. In general, fitness depends on many traits, which are affected by mutations. In a rapidly changing environment, populations are far from any fitness optimum, with many mutations available that increase fitness (and even more that decrease fitness).

To model such scenarios, we consider a collection of N asexual individuals that are characterized by a log-fitness y, which determines their average reproductive success. Specifically, the number offspring of an individual is Poisson-distributed with mean $\exp(y - \lambda)$, where $\lambda = \bar{y} - 1 + N/N_0$ keeps the population size roughly at N_0. The log-fitness of individuals is changed by mutation with probability μ per generation, where the mutational effect, δ, is drawn from a distribution $K(\delta)$. The balance between frequent mutation and selection results in a population that behaves as a traveling pulse along the fitness axis with a steady fitness variance σ^2 (Fig. 2). Absolute fitness itself is, of
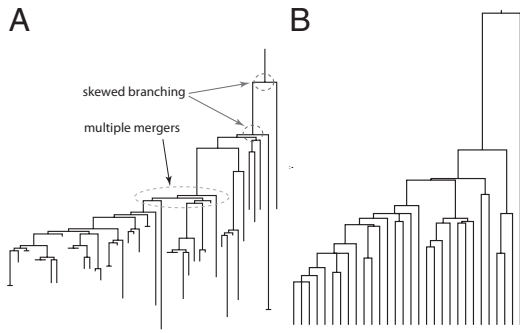
Fig. 1. *A* shows a maximum-likelihood tree of influenza nucleotide sequences (HA segment) sampled in Asia in 2009 (subtype H3N2) produced using Fasttree (13). *B* shows a tree drawn from a simulation of our model of adapting populations. Both trees often branch very unevenly, with almost all descendants on the left-most branch. Although approximate multiple mergers are common in both trees, the influenza tree does not display the uniformly long terminal branches that we observe in simulations. This could be caused by heterogeneous sampling of influenza. Trees are drawn with Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

course, not increasing indefinitely, but increasing fitness is offset by environmental deterioration and deleterious mutations.

We have implemented this model as a computer program (*SI Appendix*) that allows for different mutation distributions $K(\delta)$. In addition, the program keeps track of the parents of each new individual and thereby, saves the complete genealogy of the population. Individuals not leaving any offspring are removed from the genealogical record. From this genealogical record, quantities like pair coalescence times are readily obtained. Furthermore, we can calculate SFSs of neutral mutations by integrating over all positions in the genealogies where such mutations might have occurred.

Similar models have been used by a number of authors (22–27) who have studied the rate of adaptation in these models. Here, we focus on genealogies and their relation to observed genetic diversity. If mutations are frequent relative to the typical effect size of mutations, the model has a continuous time limit described by a stochastic differential equation for the distribution $c(y,t)$ of log-fitness y in the population (26, 28, 29)

$$\partial_t c(y,t) = D\partial_y^2 c(y,t) - \Delta_\mu \partial_y c(y,t) + (y-\lambda)c(y,t) + \text{drift}, \quad [1]$$

where the last term represents the stochastic nature of reproduction (derivation in *SI Appendix*). The diffusion constant and the average mutation input are given by $D = \mu\langle\delta^2\rangle/2$ and $\Delta_\mu = \mu\langle\delta\rangle$, respectively, where the average $\langle\ldots\rangle$ is over the distribution of mutational effects $K(\delta)$. The exact form of the distribution of mutational effects and the relative importance of deleterious and beneficial mutations are irrelevant as long as this diffusive approximation is valid (*SI Appendix*). Unless otherwise stated, we use $\mu = 1$ and draw mutational effects from a Gaussian distribution with variance s^2 and zero mean.

In this model, large populations attain a steady fitness distribution of roughly Gaussian shape with variance $\sigma^2 \approx D^{\frac{2}{3}}(24 \log \tilde{N})^{\frac{1}{3}}$, where $\tilde{N} = N D^{\frac{1}{3}}$ (22, 28). The distribution translates to higher fitness with a velocity $v = \sigma^2 + \Delta_\mu$. The distribution and its landmarks are sketched in Fig. 2. It is convenient to measure log-fitness relative to the population mean, $x = y - \bar{y}$. The fittest individuals of the population reside roughly $x_c \approx \sigma^4/4D$ above the population mean. Computer programs and analysis scripts are available on the authors' Web site (http://www.eb.tuebingen.mpg.de/research/research-groups/richard-neher.html).

Results

We first present simulation results of our model and contrast the patterns of genetic diversity of continuously adapting popula-

tions with neutral expectations. Below, we will analyze our model mathematically and show that the striking differences result from the exponential amplification of individual lineages by selection.

Distribution of Heterozygosity and Pair Coalescence Times. Assuming a molecular clock, the expected number of neutral differences between two genomes is $\pi = 2T_2\mu_n$, where μ_n is the neutral mutation rate and T_2 is the time to the most recent ancestor of the pair of sequences. Across many realizations of the process (e.g., independent loci), T_2 follows a distribution $P(T_2)$, which in the neutral case, is exponential with mean N. Simulation results for our model shown in Fig. 3 display a very different distribution of T_2 and equivalently, π. Very few pairs of sequences coalesce early, which results in the long terminal branches observed in trees (Fig. 1B). We then observe a peak in coalescence around $t \approx \sigma^2/2D$, after which the distribution of pair coalescence times decays exponentially with a characteristic time constant proportional to $\sigma^2/2D$. Within a neutral coalescent framework, a distribution of this kind would be interpreted as a rapid population expansion starting $\sigma^2/2D$ in the past. Before this expansion, the population size would be estimated to have been constant at $N_e \propto \sigma^2/2D$. However, the size of the population did not change in our model. Instead, the population was adapting by many small steps, and the conclusion that N increased in the past is wrong.

Two lineages chosen at random from the population are most likely from near the center of the fitness distribution. There are many individuals in this part of the distribution, and therefore, the probability of immediate coalescence is low. Although the sampled individuals are typical, their ancestors tend to have higher than average fitness. Only after ancestral lineages have moved to the high fitness tail of the distribution, where only few individuals are, does the rate of coalescence become appreciable. This migration of lineages to higher fitness is a well-known effect (6, 30, 31), and it is illustrated in Fig. 2. The speed at which lineages move to higher (relative) fitness is initially σ^2 (the speed of the mean minus the mutational input), whereas they slow down as they reach the tip. Consistent with the above interpretation, the delay of coalescence, T_{delay}, is roughly two times the time required for the mean fitness to catch up with the high fitness nose (i.e., $T_{delay} \approx 2x_c\sigma^{-2} = \sigma^2/2D$).
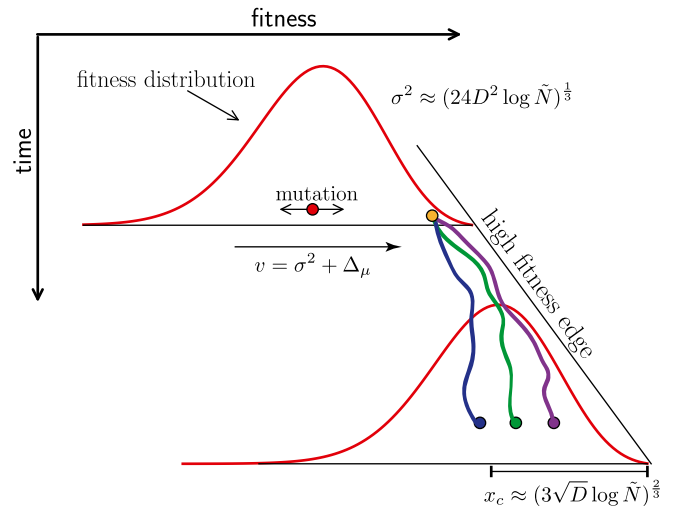
Fig. 2. Ancestral lineages in evolving populations. The figure shows the fitness distribution of the population, translating to higher fitness with velocity $\sigma^2 + \Delta_\mu$, at two time points. Randomly sampled individuals (green, blue, and violet dots in the later population) tend to come from the center of the distribution, whereas ancestors tend to be among the fittest in the population. The ancestral lineages wiggle because of mutations that randomly perturb their fitness. Simultaneously, lineages move to the high fitness edge, where they are likely to meet and coalesce. The fittest individuals are typically at $x_c \approx \sigma^4/4D$ above the mean fitness.

Neher and Hallatschek
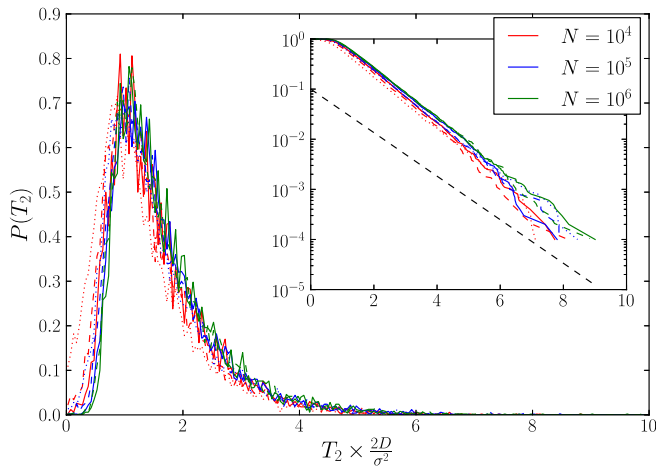
Fig. 3. The distribution of pair coalescence times (proportional to heterozygosity) in a model of rapidly adapting populations. After rescaling time by $\sigma^2/2D$, curves for different N and s collapse onto a single master curve. This collapse shows that $\sigma^2/2D \propto D^{-\frac{1}{3}}(\log \bar{N})^{\frac{1}{3}}$ is the timescale of coalescence. After a delay, $T_{delay} \approx \sigma^2/2D$, T_2 is exponentially distributed, which is apparent from the *Inset* showing the cumulative distribution $P(T_2 > T)$. An exponential $\exp[-2TD/\sigma^2]$ is indicated as a black dashed line. Different line styles correspond to $s = 0.01$ (solid), $s = 0.001$ (dashed), and $s = 0.0001$ (dotted), whereas the mutation rate is $\mu = 1$. For each parameter combination, random pairs are sampled at 10,000 time points $2s^{-2/3}$ generations apart. Fig. S1 shows the corresponding distributions of T_3 and T_4.

After lineages have moved to the high fitness tail, they seem to coalesce uniformly with a time constant $T_c \approx \sigma^2/2D$. From the dependence of σ^2 on population parameters, we see that $T_c \propto (\log \bar{N})^{\frac{1}{3}}$ increases only weakly with the population size.

Site Frequency Spectra. The density $f(\nu)d\nu$ of neutral-derived alleles in the frequency interval $[\nu, \nu + d\nu]$ is known as the SFS. The neutral SFS is a convenient summary of the neutral diversity segregating in the population. A mutation that happened on a particular branch of the genealogy will later be present in all individuals that descend from this branch. Hence, the SFS harbors information about the distribution of branch weights and the branch length of the genealogy. In Kingman's coalescent, the SFS is simply given by $f(\nu) = \Theta/\nu$, where $\Theta = 2\langle T_2 \rangle \mu$ is the average heterozygosity. Importantly, it is a monotonically decreasing function of the frequency. Fig. 4 shows SFSs measured in simulations of our model. The most striking qualitative difference is the nonmonotonicity, a feature known to be common in the presence of selective sweeps caused by hitchhiking (32).

The nonmonotonicity of $f(\nu)$ implies the existence of long branches deep in the tree that are ancestors of almost everybody in the population, whereas a small minority of the population descends from different lineages. Such very asymmetric branchings are unlikely in Kingman's coalescent, where at any split, the fraction of individuals that go left or right is uniformly distributed (2). Such asymmetric branchings are common in our model and frequently observed in reconstructed genealogies from rapidly adapting organisms (Fig. 1A).

The axes in Fig. 4 are scaled to facilitate the comparison with analytic results. At low frequencies, the SFS is proportional to ν^{-2} (33, 34) and hence much steeper than the neutral SFS in Kingman's coalescent, $f(\nu) \propto \nu^{-1}$. Hence, samples will be dominated by singletons. In addition, $f(\nu)$ is nonmonotonic and increases as $\nu \to 1$.

The majority of the contributions to the increase of $f(\nu)$ for $\nu \to 1$ stems from the very last coalescent event. In this last coalescent event, two or more lineages are merging. One of these lineages is typically the ancestor of almost the entire sample, whereas the others share the remaining minority. The distribution of the offspring of these lineages and their number has been studied by

Goldschmidt and Martin (35), who showed that the distribution of the size of the biggest lineage is asymptotically $\propto (1-\nu)^{-1}$. In *SI Appendix*, section III, we derive the more accurate approximation

$$f(\nu) \approx \frac{T_c \mu}{(\nu - 1)\log(1 - \nu)} \quad \text{for} \quad 1 - \nu \ll 1. \quad \text{[2]}$$

To compare the SFS of our model with SFS of the BSC across the entire range of ν, we simulated the idealized BSC and find very good agreement (solid black line in Fig. 4). The SFS of the idealized BSC deviates from the SFS of the model of adaptation only at very low allele frequencies. The model of adaptation tends to have even more rare alleles than the BSC, which is because of the fact that lineages have to move to the high fitness tail before coalescence begins.

The nonmonotonicity $f(\nu)$ is a clear indication that the genealogies in this model with selection are fundamentally different from canonical neutral genealogies (Kingman). In Kingman's coalescent, neither constant nor exponentially growing population sizes gives rise to nonmonotonic SFS (Fig. S2).

Time to the Most Recent Common Ancestor. In Kingman's coalescent, the expected time to the most recent common ancestor (MRCA) of a sample of size n, $\langle T_{MRCA} \rangle = N(2 - 2/n)$, increases only very slowly with n. This saturation is a consequence of the even branching ratios; an additional individual will most likely coalesce with existing samples and only rarely increase T_{MRCA}. In contrast, the trees generated by our model of adaptation tend to branch very unevenly, and one often observes that one external branch goes all of the way back to the MRCA of the sample, as in Fig. 1B. As the sample size is increased, one continues to sample deeper into the tree. This increasing tree depth is a generic property of the BSC, where the average $\langle T_{MRCA} \rangle$ increases as $\log \log n$ with the sample size n. Similarly, $\langle T_{MRCA} \rangle$ of the entire population is expected to increase as $T_c \log \log N$ with the population size. Our simulations are consistent with this behavior (Fig. S3).

Note that T_c depends weakly on N in adapting populations, whereas it increases linearly with N in Kingman's coalescent. In contrast, the rescaled time to the MRCA, $T_c^{-1}\langle T_{MRCA} \rangle$, asymptotes to two in Kingman's coalescent, whereas it continues to increase with N in the BSC.
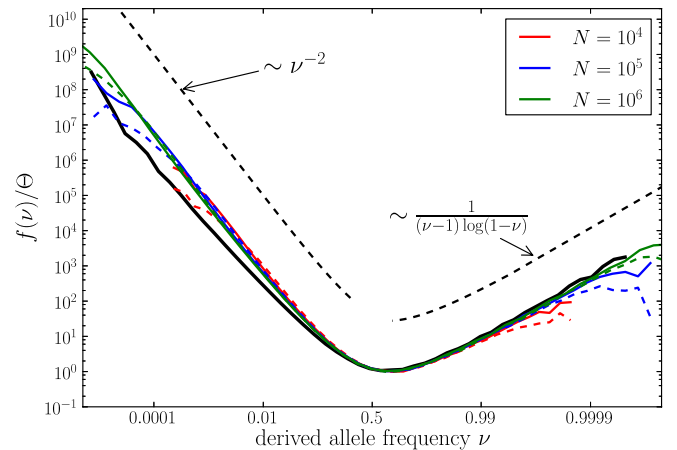
Fig. 4. SFS of (derived) neutral alleles in rapidly adapting populations is nonmonotonic, with peaks at low frequencies and near fixation. The asymptotic behavior of the SFS at low and high derived frequencies is shown as dashed black lines. The solid black line is the SFS of the BSC simulated using Eqs. **7** and **8** with $N = 100,000$ averaged over 10,000 runs. Line styles and parameters are as in Fig. 3.

Analysis. The simulation results presented above show that genealogies arising in our model are distinct from those genealogies expected in Kingman's coalescent and display a number of features reminiscent of the BSC. We will now describe how this coalescent process emerges from the dynamics of the model.

Individuals in our model have a heritable fitness that determines the distribution of the number of immediate offspring. Although fit individuals have, on average, more offspring than less-fit individuals, the fitness differences in the population are small, and the offspring distribution across the population is narrow. However, fitness is heritable, and fit individuals can have a very large number of distant great^t-grandchildren. Hence, the distribution of offspring after t generations, $P(n,t)$, will be dominated by fit individuals and can have a very long tail. Conversely, the present-day population has fewer and fewer ancestors as we trace its lineages back in time. At T_{MRCA} generations in the past, there is exactly one individual that is the ancestor to the entire population. Ancestors of the MRCA are also common ancestors (CAs) of the entire population, albeit not the most recent one. Fig. 5 shows that MRCAs and CAs tend to come from the high fitness tail of the population. MRCAs tend to be fitter than CAs, because they are conditioned on giving rise to at least two lineages that persist to the present.

The offspring distribution, $P(n,t)$, changes slowly from the initial narrow distribution to a broad distribution with a power law at intermediate times (34). The broad distribution at intermediate times is at the heart of the correspondence of genealogies in models of adapting populations and the BSC.

The BSC assumes that all individuals are exchangeable and that, in every coalescent event, a randomly chosen set of lineages merges into one. Each possible merger event has a specific rate associated with it, and the rate at which k individuals merge into one CA is $q_k = T_2^{-1}(k-1)^{-1}$ (16) (the general expression for the rates is given below in Eq. **7**). In contrast, in the neutral coalescent, higher-order coalescence is very rare, $\propto N^{-(k-1)}$. We will present the basic properties of the BSC briefly in the discussion.

To appreciate how these coalescence rates can emerge from a model with selection, consider the number of individuals n_i that descend from an individual i that lived t generations in the past. The probability that k individuals sampled randomly from the population have a CA t in the past is then given by (Eq. **3**)

$$Q_k(t) = \left\langle \sum_{i=1}^{N} \left(\frac{n_i}{\sum_j n_j} \right)^k \right\rangle, \qquad [3]$$

where the average $\langle . \rangle$ is over all n_i. $Q_k(t)$ is dominated by $n_i \gg k$, and therefore, sampling with replacement in Eq. **3** is an accurate

approximation. Using the identity $\Gamma(k)C^{-k} = \int_0^z dz\, z^{k-1} e^{-zC}$ and assuming that t is small enough that the different n_i values are still approximately independent, we can express Q_k as

$$Q_k(t) \approx \frac{N}{\Gamma(k)} \int_0^\infty dz\, z^{k-1} \langle e^{-zn} \rangle^{N-1} \langle n^k e^{-zn} \rangle$$

$$\approx -\frac{N}{\Gamma(k)} \int_0^\infty dz\, z^{k-1} e^{-N\Phi_z(t)} (-1)^k \partial_z^k \Phi_z(t), \qquad [4]$$

where we introduced the Laplace transform $1 - \Phi_z(t) = \sum_n e^{-zn} P(n,t)$ and assumed $N\Phi_z^2(t) \ll 1$. In *SI Appendix*, we show that $\Phi(t) \sim z^{\frac{\sigma^2}{2Dt}}$ for $t > T_{delay} = \sigma^2/2D$. For a limited interval after $t > T_{delay}$, we find that the probability that k individuals have a CA increases with rate

$$q_k = \frac{2D}{\sigma^2} \frac{1}{k-1} \qquad [5]$$

per unit time. More general coalescence rates can be calculated analogously (*SI Appendix*). Before T_{delay}, the rate of coalescence is very low. This result is in agreement with Fig. 3, where we found that little coalescence happened early, whereas coalescence times are exponentially distributed after that time, with characteristic time $T_c \approx \sigma^2/2D$ for $t > T_{delay}$. The relative rates of mergers of two, three, etc. are consistent with the BSC, explaining our observations for the frequency spectrum and the time to the MRCA.

The branching process approximation used to derive the result in Eq. **5** is valid only for short times but nevertheless, gives us the relative rates of multiple mergers after coalescence begins. For subsequent deeper coalescent events, the relevant lineages are already at the tip of the fitness distribution, and this process repeats itself without the delay. In fact, after this delay, all remaining lineages are in a narrow region at the tip of the fitness distribution. The situation now resembles the situation of coalescence in FKPP waves: The fitness of the lineages is roughly equal, but lineages have to stay ahead of a fitness cutoff to survive. We can, therefore, use the phenomenological theory of genealogies in FKPP waves from ref. 36, which confirms the above result for the coalescent timescale (*SI Appendix*). In *SI Appendix*, we present an additional argument based on tuned models introduced in ref. 26.
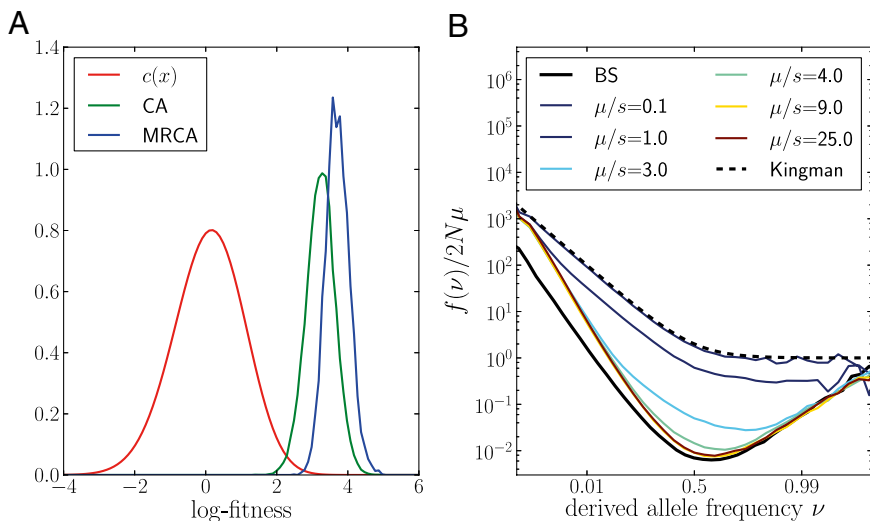
Fig. 5. *A* shows the distribution of the log-fitness of all CAs and all MRCAs compared with the average distribution, $c(x)$, of log-fitness in the population (described in the text). These distributions were measured in forward simulations with $\mu = 1, s = 0.01$, and $N = 10^6$. (*B*) SFS of derived neutral alleles in a background selection scenario with deleterious mutations of effect s. As the ratio μ/s is varied while keeping $\sigma^2 = \mu s = 0.01$ constant, the SFS interpolates between the expectation for the Kingman's coalescent and the BSC ($N = 10^4$).

To corroborate our analysis, we performed additional simulations that allow us to measure the Laplace transform of the distribution of pair coalescent times for very large populations. These simulations show that the pair coalescent time is, indeed, exponential with characteristic time $\sigma^2/2D$ after a delay of the same length (Fig. S4). The algorithm used is similar in spirit to the algorithm by Brunet et al. (18) (*SI Appendix*).

Strictly speaking, the analogy to an exchangeable coalescent model like the BSC requires that different coalescence events that one lineage undergoes be independent. For this finding to be true, individuals descending from a lineage have to distribute evenly across the fitness distribution $c(x)$ between coalescence events, which requires a time $T_{eq} \approx T_c$. Hence, we should not expect a clean convergence to the BSC. Nevertheless, we find it to be a very good model for the observed genealogies after accounting for the delay. The underlying reason is that local equilibration in the region where the ancestral lineages are is fast ($t \approx D^{-1/3}$). This region, however, undergoes fluctuations on the timescale T_{eq} that modulate the overall rate of coalescence but do not significantly affect the local dynamics. For waves of FKPP type that describe the spread of individuals in space, $T_{eq} \ll T_c$ in large populations (18).

Discussion

We have shown that, in a simple model of adapting populations, the observed genealogies are inconsistent with the standard neutral coalescent. Instead, genealogical trees are characterized by long terminal branches and almost simultaneous coalescence of multiple lineages. At branching events deep in the tree, one commonly observes that almost all individuals of the population descend from one branch, whereas very few descend from the other branches. Such skewed branching is unlikely in standard neutral coalescent models, regardless of the history of the effective population size. One consequence of these uneven branching ratios is a nonmonotonic SFS of derived neutral alleles. Compared with the Kingman's coalescent, the low-frequency part of the SFS is much steeper, whereas the high-frequency part shows a characteristic upturn (Fig. 4).

A given pair of lineages is unlikely to coalesce in the bulk of the fitness distribution. Typically, both lineages move into the high-fitness tip of the population distribution before they coalesce as illustrated in Fig. 2. These dynamics result in long terminal branches and a distribution of heterozygosities peaked at intermediate values. After this delay, the typical time to coalescence is again on the order of the time that it takes the fittest individuals to dominate the population (Fig. 3). In panmictic populations, this time depends on the logarithm of the population size, and in our model, it is proportional to $(\log \tilde{N})^3$.

We argue that the exponential amplification of fit lineages is responsible for these observations and that coalescence in such rapidly adapting populations is generically described by a modified BSC (16, 17). The BSC is a special case of the large class of Λ-coalescent processes (15). Given the distribution $p(f)$ of the fraction f of the population that descends from a single individual in the previous generation, the rate at which k of b lineages merges is given by

$$\lambda_{b,k} = \int df\, p(f) f^k (1-f)^{b-k}. \qquad [6]$$

The BSC corresponds to $p(f) \sim f^{-2}$ for large f, in which case Eq. 6 reduces to

$$\lambda_{b,k} = \frac{1}{T_c} \frac{(k-2)!(b-k)!}{(b-1)!} \qquad [7]$$

or Eq. 5 for the special case $b=k$. The total rate at which coalescence events happen in a sample of k lineages is, therefore,

$$\lambda_b = \sum_k \binom{b}{k} \lambda_{b,k} = \frac{b-1}{T_c} \qquad [8]$$

in contrast to the neutral coalescent, where $\lambda_b \propto b(b-1)$. A coalescence event reduces the number of surviving lineages on average by $\log b$, and therefore, the average rate at which the number of lineages decreases is $\approx T_c^{-1} b \log b$. The typical time needed to reach the CA of a sample of size n is $\approx T_c \log \log n$, in contrast to $2T_c$ in Kingman's coalescent. The BSC occupies a special intermediate position between Kingman's coalescent, where only pairwise mergers are allowed, and a star coalescent, where all lineages coalesce simultaneously. Star-like genealogies are expected in rapidly expanding populations or a region fully linked to a recent rapid hard sweep (37). In the BSC, multiple mergers (subsets of lineages with star-like trees) are frequent, but at the same time, there are many mergers at different depths in the tree. In fact, the BSC is the $\alpha = 1$ case of the one-parameter family of β-coalescents with parameter $0 < \alpha < 2$, whereas the case $\alpha \to 2$ corresponds to the Kingman's coalescent. A more in-depth discussion is in the recent review by Berestycki (16). The BSC is easily implemented as a computer simulation by drawing an exponentially distributed random number with mean λ_b^{-1} to determine the time of the next event. The type of event is then chosen with probabilities proportional to $\lambda_{b,k}$.

The models of adaptation that we have studied have a narrow offspring distribution. Nevertheless, the exponential amplification of fit genotypes over many generations gives rise to a distribution of clone sizes with the required asymptotic behavior. The important lineages are those lineages that run ahead of the distribution, expand faster, and take over a significant fraction of the population (36). Over even longer times, the fitness of ancestors and descendants decorrelates. This gradual decorrelation allows us to approximate the genealogies with the abstract BSC, which assumes that there are no correlations in offspring number across generations.

Conventionally, an increased variance in offspring number is accounted for by defining an effective population size. With a clone size distribution $p(f) \sim f^{-2}$, however, the variance diverges with the population size (34). Similar effects arise in other models with very skewed offspring distribution (38). As a consequence, the genealogies are dominated by rare anomalously large clones and described by the BSC rather than Kingman's coalescent. The rate of coalescence is not set by N^{-1} but by the rate at which clones expand and collapse. We would like to stress that evolutionary dynamics thereby remain highly stochastic, even in very large populations. Analogous behavior has recently been observed in models of individuals invading uninhabited territory (FKPP type waves) (18) and ensembles of supercritical branching processes (39).

The BSC is not only a good model for genealogies of adapting asexual populations, but also, it applies to populations under purifying selection in which Muller's ratchet clicks often. The standard model for the distortion of genealogies by purifying selection assumes that deleterious mutations are rapidly purged and coalescence is neutral in the mutation-free class with a reduced population size $Ne^{-\mu/s}$, where μ is the deleterious mutation rate and s is the effect size of deleterious mutations (4). More elaborate analysis based on a fitness class coalescent explicitly tracks lineages through the population and calculates the contribution to coalescence before lineages reach the mutation-free class (5). However, this standard model of purifying selection only applies if the mutation-free class is large and Muller's ratchet does not operate, which requires $Nse^{-\mu/s} \ll 1$ (40–42). Fig. 5B shows the SFS of derived neutral alleles for different ratios μ/s. For small μ/s, the SFS is similar to the SFS of Kingman's coalescent with a reduced time to coalescence, in accordance with the background selection theory. However, as soon as the ratchet starts to click frequently, the SFS develops the nonmonotonicity characteristic of the BSC.

If the ratchet is clicking fast, the fitness distribution in the population resembles the distribution of traveling wave models,

but selection on fitness variation cannot keep up with the influx of deleterious mutations. Similarly, populations in a steady balance between deleterious and beneficial mutations (43) have genealogies as found here for rapidly adapting populations. The reason for the qualitative difference in the ratchet regime is the fact that the nose of the wave is not steady but constantly turning over. Different lineages are struggling to get ahead of everybody else and in the frame of reference of the population (that is, relative to mean fitness), exponentially amplified. In contrast, dynamics of lineages in the mutation-free class are neutral if Muller's ratchet does not operate.

In *SI Appendix*, we show that the argument that gave rise to the particular coalescence rates in Eq. **5** can be extended to a large class of models that are controlled by a small and fluctuating population of highly fit individuals. We argue that the BSC generically emerges as a consequence of the exponential amplification of the clones descending from these highly fit individuals together with the seeding of novel lineages. The latter could happen by lucky diffusion to high fitness (our model), large-effect beneficial mutations, or lucky outcrossing. After some time, the distribution of lineage size follows a power law with an exponent close to −2 (23, 34, 44). Given an effective offspring distribution of this shape, the BSC follows (18, 39). In ref. 45, the authors study a model where the mutation rate is much smaller than the typical effect sizes of mutations. They show that, also in this case, the genealogies are well-approximated by the BSC after a delay (ref. 45; see also Figs. S5 and S6). Whether the BSC also describes genealogies in scenarios where fitness is increased in rather large increments (compared with the population diversity) (46, 47) remains an interesting topic for future work.

The compatibility of a sample with the neutral coalescent model is typically assessed using statistics such as Tajima's D (48). Tajima's D compares the average number of pairwise differences with the total number of segregating sites in the sample. In the case of the BSC, the average pairwise diversity is proportional to $\langle T_2 \rangle$, whereas the total number of segregating sites is proportional to $\langle T_2 \rangle n / \log n$ (compared with $\langle T_2 \rangle \log n$ for the Kingman's coalescent). This tremendous excess of segregating sites is a consequence of the very steep SFS at small frequencies and results in $D \propto -\log(n)$.

Sexual populations and recurrent selective sweeps at linked loci can also give rise to multiple mergers in the genealogies (9). However, recombination and sexual reproduction will reduce the effects of linked selection and decouple the genealogies of different loci. Hence, we expect that the coalescent behavior crosses over to Kingman's coalescent as the recombination rate increases—at least in models of panmictic populations. This crossover is observed in models of facultatively sexual populations (34).

Given the apparent universality of the BSC in spatially expanding populations and panmictic adapting populations, it should be included as a prior in popular population genetic and phylogenetic inference programs such as BEAST (49).

1. Kingman J (1982) On the genealogy of large populations. *J Appl Probab* 19A:27–43.
2. Derrida B, Peliti L (1991) Evolution in a flat fitness landscape. *Bull Math Biol* 53(3):355–382.
3. Nordborg M (1997) Structured coalescent processes on different time scales. *Genetics* 146(4):1501–1514.
4. Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.
5. Walczak AM, Nicolaisen LE, Plotkin JB, Desai MM (2012) The structure of genealogies in the presence of purifying selection: A fitness-class coalescent. *Genetics* 190(2):753–779.
6. O'Fallon BD, Seger J, Adler FR (2010) A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. *Mol Biol Evol* 27(5):1162–1172.
7. Barton NH, Etheridge AM (2004) The effect of selection on genealogies. *Genetics* 166 (2):1115–1131.
8. Barton N (1998) The effect of hitch-hiking on neutral genealogies. *Genet Res* 72(2): 123–133.
9. Durrett R, Schweinsberg J (2005) A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stochastic Process Appl* 115(10):1628–1657.
10. Krone SM, Neuhauser C (1997) Ancestral processes with selection. *Theor Popul Biol* 51(3):210–237.
11. Bedford T, Cobey S, Pascual M (2011) Strength and tempo of selection revealed in viral gene genealogies. *BMC Evol Biol* 11:220.
12. Seger J, et al. (2010) Gene genealogies strongly distorted by weakly interfering mutations in constant environments. *Genetics* 184(2):529–545.
13. Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26(7):1641–1650.
14. Strelkowa N, Lässig M (2012) Clonal interference in the evolution of influenza. *Genetics* 192(2):671–682.
15. Pitman J (1999) Coalescents with multiple collisions. *Ann Probab* 27(4):1870–1902.
16. Berestycki N (2009) Recent progress in coalescent theory. arXiv:math.PR/0909.3985.
17. Bolthausen E, Sznitman A-S (1998) On Ruelle's probability cascades and an abstract cavity method. *Commun Math Phys* 197(2):247–276.
18. Brunet E, Derrida B, Mueller AH, Munier S (2007) Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Phys Rev E Stat Nonlin Soft Matter Phys* 76(4 Pt 1):041104.
19. Fisher RA (1937) The wave of advance of advantageous genes. *Ann Eugen* 7(4):355–369.
20. Kolmogorov A, Petrovskii I, Piscounov N (1937) Etude de l'equation de la diffusion avec croissance de la quantite de matiere et son application a un probleme biologique. *Bull Moscow Univ Math Mech* 1:1–25.
21. Brunet É, Derrida B (2012) Genealogies in simple models of evolution. arXiv:q-bio.PE.
22. Tsimring LS, Levine H, Kessler DA (1996) RNA virus evolution via a fitness-space model. *Phys Rev Lett* 76(23):4440–4443.
23. Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176(3):1759–1798.
24. Rouzine IM, Wakeley J, Coffin JM (2003) The solitary wave of asexual evolution. *Proc Natl Acad Sci USA* 100(2):587–592.
25. Neher RA, Shraiman BI, Fisher DS (2010) Rate of adaptation in large sexual populations. *Genetics* 184(2):467–481.
26. Hallatschek O (2011) The noisy edge of traveling waves. *Proc Natl Acad Sci USA* 108 (5):1783–1787.
27. Park SC, Krug J (2007) Clonal interference in large populations. *Proc Natl Acad Sci USA* 104(46):18135–18140.
28. Cohen E, Kessler DA, Levine H (2005) Front propagation up a reaction rate gradient. *Phys Rev E Stat Nonlin Soft Matter Phys* 72(6 Pt 2):066126.
29. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM (2012) Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc Natl Acad Sci USA* 109(13):4950–4955.
30. Hermisson J, Redner O, Wagner H, Baake E (2002) Mutation-selection balance: Ancestry, load, and maximum principle. *Theor Popul Biol* 62(1):9–46.
31. Rouzine IM, Coffin JM (2007) Highly fit ancestors of a partly sexual haploid population. *Theor Popul Biol* 71(2):239–250.
32. Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155(3): 1405–1413.
33. Basdevant A-L, Goldschmidt C (2008) Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electron J Probab* 13(17):486–512.
34. Neher RA, Shraiman BI (2011) Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* 188(4):975–996.
35. Goldschmidt C, Martin JB (2005) Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron J Probab* 10(21):718–745.
36. Brunet E, Derrida B, Mueller AH, Munier S (2006) Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys Rev E Stat Nonlin Soft Matter Phys* 73(5 Pt 2):056126.
37. Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129(2):555–562.
38. Eldon B, Wakeley J (2006) Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* 172(4):2621–2633.
39. Schweinsberg J (2003) Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process Appl* 106(1):107–139.
40. Stephan W, Chao L, Smale JG (1993) The advance of Muller's ratchet in a haploid asexual population: Approximate solutions based on diffusion theory. *Genet Res* 61(3):225–231.
41. Jain K (2008) Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics* 179(4):2125–2134.
42. Neher RA, Shraiman BI (2012) Fluctuations of fitness distributions and the rate of Muller's ratchet. *Genetics* 191(4):1283–1293.
43. Goyal S, et al. (2012) Dynamic mutation-selection balance as an evolutionary attractor. *Genetics* 191(4):1309–1319.
44. Yule GU (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos Trans R Soc Lond B Biol Sci* 213:21–87.
45. Desai MM, Walczak AM, Fisher DS (2012) Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 10.1534/genetics.112.147157.
46. Gerrish PJ, Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103(1-6):127–144.
47. Schiffels S, Szöllosi GJ, Mustonen V, Lässig M (2011) Emergent neutrality in adaptive asexual evolution. *Genetics* 189(4):1361–1375.
48. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
49. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.

Supplementary Information: Genealogies of rapidly adapting populations

Richard A. Neher

Evolutionary Dynamics and Biophysics Group, Max Planck Institute for Developmental Biology, 72076, Tübingen, Germany

Oskar Hallatschek

Biophysics and Evolutionary Dynamics Group, Max Planck Institute
for Dynamics and Self-Organization, 37077 Göttingen, Germany
(Dated: December 14, 2012)

I. DERIVATION AND LIMITATIONS OF EQ. 1

We assume a population of N individuals that replicate at rate B, die at rate D, and mutate at rate μ. For convenience, we measure time in units of generations and set birth and death rate to $B = \exp(y - \lambda)$ and $D = 1$, respectively. Here, y is the log-fitness of the individual in question, while the population size is kept constant by adjusting λ. Typically, λ is equal to the sum of the mean log-fitness and a small modulation that increases growth rate when the population size falls below the target N_0 and decreases growth rate otherwise. We use $\lambda = \bar{y} - (1 - NN_0^{-1})$ which keeps N within $\mathcal{O}(\sqrt{N_0})$ of N_0. For present purposes, this model is equivalent to a model with strictly constant population size. Different forms of λ that constrain the population size to a different degree are possible [1]. Mutations increment the log-fitness y of an individual by δ drawn from a distribution $K(\delta)$. Disregarding fluctuations and assuming fitness differences are small, the distribution of fitness in the population, $c(y,t)$, obeys the following deterministic equation

$$\partial_t c(y,t) = (y - \lambda)c(y) + \mu \int ds\, K(\delta)\left[c(y - \delta, t) - c(y,t)\right] \tag{1}$$

Generically, steady state solutions to this equation oscillate around 0 for large $y - \bar{y}$, where the stochastic effects of finite populations become important [2]. The magnitude of stochastic perturbations to the population in the interval $[y, y + \Delta y]$ are proportional to $\sqrt{c(y)\Delta y}$ and can be accounted for either by an explicit model of discrete particles, or a stochastic partial-differential equation

$$\partial_t c(y,t) = (y - \lambda)c(y) + \mu \int d\delta\, K(\delta)\left[c(y - \delta, t) - c(y,t)\right] + \sqrt{c(y,t)}\eta(y,t) \tag{2}$$

where $\eta(y,t)$ is Gaussian white noise with $\langle \eta(y,t)\eta(y+x,t')\rangle = \delta(x)\delta(t - t')$ [1]. This stochasticity determines the average steady state velocity, which diverges in the limit $N \to \infty$ [2–4].

If the mutation rate μ is large and the distribution $K(\delta)$ falls off quickly enough such that the typical mutations are small, one can expand $c(y - \delta, t)$ inside the integral to obtain

$$\partial_t c(y,t) \approx (y - \lambda)c(y,t) + \frac{\mu\langle\delta^2\rangle}{2}\frac{\partial^2}{\partial y^2}c(y,t) - \mu\langle\delta\rangle\frac{\partial}{\partial y}c(y,t) + \sqrt{c(y,t)}\eta(y,t) \tag{3}$$

where $\langle\ldots\rangle$ denotes the average over $K(\delta)$. The expansion of $c(y - \delta, t)$ is a good approximation as long as the mutation rate is larger than typical mutational effects. Specifically, it is required that $\mu^2/\langle\delta^2\rangle > (\log \tilde{N})^2$, see [2] for a discussion of this issue. This condition will limit the applicability of the model to RNA virus populations or asexual eukaryotes with long genomes. We find, however, that the basic features of the BSC are also present in models where $\mu/\sqrt{\langle\delta^2\rangle} \ll 1$, see supplementary figures S5 and S6 below. This observation is confirmed by a recent preprint by Desai et al [5], in which the authors investigate a model with $\mu^2 \ll \langle\delta^2\rangle$.

The average of the population distribution governed by Eq. (3) obtains a steady shape that translates towards higher y with a velocity $v = \sigma^2 + \mu\langle\delta\rangle$, where

$$\sigma^2 \approx D^{\frac{2}{3}}\left(24\log(ND^{\frac{1}{3}})\right)^{\frac{1}{3}} \tag{4}$$

and $D = \frac{\mu\langle\delta^2\rangle}{2}$. In a suitably chosen comoving frame $x = y - vt$, the equation describing the population distribution simplifies to

$$0 \approx xc(x,t) + D\frac{\partial^2}{\partial x^2}c(y,t) + \sigma^2\frac{\partial}{\partial x}c(x,t) + \sqrt{c(x,t)}\eta(x,t) \tag{5}$$

If the mutational input $\mu\langle\delta\rangle = 0$, the average velocity equals the fitness variance σ^2 of the population. If deleterious mutations are more common than beneficial mutations and on average reduce fitness, the velocity will be smaller than the fitness variance by $|\mu\langle\delta\rangle|$. Eq. (5) has been studied in detail in [2]. In large population, the typical solution in the quasi-steady state is given by

$$c(x) = \begin{cases} Ce^{-\frac{\sigma^2 x}{2D}}\text{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x}{D^{\frac{1}{3}}}\right) & x < x_c \\ De^{-\frac{\sigma^2 x}{D}} & x > x_c \end{cases} \tag{6}$$

where x_c and D are determined by matching $c(x)$ and $\partial_x c(x)$ at x_c, while C is set by the normalization.

II.   FREQUENCY SPECTRUM OF DERIVED ALLELES IN THE BSC

A mutation occurring on some branch of the genealogy will be present in all individuals that descend from the branch considered. Very recent mutations tend to be rare, while old mutations can be common. The number of leafs descending from lineages at a time $\tau = T/T_c$ in the past is also known as the size of the τ-families (T_c is the coalescent time scale). These family sizes follow a two parameter Poisson-Dirichlet distribution $PD(0, e^{-\tau})$ and the size z of the first block is Beta distributed as [6]

$$P(z,\tau) = \frac{z^{-e^{-\tau}}(1-z)^{e^{-\tau}-1}}{\Gamma(1-e^{-\tau})\Gamma(e^{-\tau})} \tag{7}$$

Common mutations, i.e., those present in the majority of the population, most likely fall onto this block. The present day frequency of the derived allele is equal to the block size. If neutral mutations occur at rate μ, we can calculate the ensemble averaged site frequency spectrum (SFS) by integrating over all times at which the mutation could have occurred. Ignoring mutations that fall onto other blocks of the Poisson-Dirichlet distribution, the average SFS is given by

$$\begin{aligned} f(\nu) &= \mu \int_0^\infty dT \frac{\nu^{-e^{-\tau}}(1-\nu)^{e^{-\tau}-1}}{\Gamma(1-e^{-\tau})\Gamma(e^{-\tau})} \\ &= \mu T_c \int_0^\infty d\tau \frac{\nu^{-e^{-\tau}}(1-\nu)^{e^{-\tau}-1}}{\Gamma(1-e^{-\tau})\Gamma(e^{-\tau})} \\ &= \mu T_c \int_0^1 dy \frac{\nu^{-y}(1-\nu)^{y-1}}{y\Gamma(1-y)\Gamma(y)} \\ &= \mu T_c \int_0^1 dy \frac{\sin(\pi y)\nu^{-y}(1-\nu)^{y-1}}{\pi y} \end{aligned} \tag{8}$$

where we have used Euler's reflection formula for the Gamma functions. The last integral has an expression in terms of special functions. However, the term $\sin(\pi y)/\pi y$ is close to one for small y where the dominant contribution to the integral comes from. Hence we have approximately

$$f(\nu) \approx \mu T_c \int_0^1 dy \nu^{-y}(1-\nu)^{y-1} = \mu T_c \frac{2\nu - 1}{\nu(1-\nu)(\log\nu - \log(1-\nu))} \tag{9}$$

Higher order corrections to Eq. (9) can be obtained by expanding the sine to higher order. They do not change the result qualitatively. For common derived alleles with $1 - \nu \ll 1$, the above restriction to the first block of the Poisson-Dirichlet distribution is a good approximation since other blocks are almost always short $z \ll 1$. In the limit $1 - \nu \ll 1$, the expression simplifies to

$$f(\nu) \approx -\frac{\mu T_c}{(1-\nu)\log(1-\nu)} \tag{10}$$

which coincides with Eq. (2) of the main text. For $1 - \nu \ll 1$, it describes the simulation data very well, see Fig. 4. At the rare end of the spectrum, the observed power law $f(\nu) \sim \nu^{-2}$ is a (well-known) consequence of exponentially growing sub-populations [7], corresponding to expanding clones in the high-fitness tip of the traveling wave.

III. GENERATING FUNCTION OF THE OFFSPRING NUMBER DISTRIBUTION

To calculate the moments of the lineage size distribution, we need the distribution $P(n,t|x,t_0)$ of lineage sizes n at time t, given the lineage was seeded a fitness x above the mean at time t_0. $P(n,t|x,t_0)$ obeys the backward equation

$$P(n,t|x+v\Delta t, t_0 - \Delta t) = [1 - \Delta t(2+x+\mu)]\, P(n,t|x,t_0) + \Delta t(1+x)\sum_{n'} P(n',t|x,t_0)P(n-n',t|x,t_0)$$

$$+ \Delta t \delta_{n,0} + \Delta t \mu \int ds\, K(\delta)P(n,t|x+\delta,t_0) \tag{11}$$

where $K(\delta)$ is a normalized distribution of mutational effects as used in Sec. I and we have used $1+x$ as birth rate and 1 as death rate. The first term gives the probability that no birth, death, or mutation is happening in the interval Δt, the second term corresponds to birth, the third term to loss of the lineage ($n=0$) and the last term to a mutation changing the log-fitness by δ. Rearranging and taking the limit $\Delta t \to 0$ yields

$$-\partial_{t_0}P(n,t|x,t_0) + v\partial_x P(n,t|x,t_0) = -(2+x+\mu)P(n,t|x,t_0) + (1+x)\sum_{n'}P(n',t|x,t_0)P(n-n',t|x,,t_0)$$

$$+ \delta_{n,0} + \mu \int d\delta\, K(s)P(n,t|x+\delta,t_0) \tag{12}$$

Using the time translation invariance, $P(n,t|x,t_0) = P(n, t-t_0|x,0)$, and defining $\psi_z(x,t) = \sum_n e^{-zn}P(n,t|x,0)$, we find

$$\partial_t \psi_z(x,t) + v\partial_x \psi_z(x,t) = 1 - (2+x)\psi_z(x,t) + (1+x)\psi_z^2(x,t) + \mu \int ds\, K(\delta)[\psi_z(x+\delta,t) - \psi_z(x,t)] \tag{13}$$

Upon multiplying by -1 and substituting $\phi_z(x,t) = 1 - \psi_z(x,t)$, this equation simplifies to

$$\partial_t \phi_z(x,t) + v\partial_x \phi_z(x,t) = x\phi_z(x,t) - \phi_z^2(x,t) + \mu \int ds\, K(\delta)[\phi_z(x+\delta,t) - \phi_z(x,t)] \tag{14}$$

where we have approximated $(1+x)\phi_z^2$ by ϕ_z^2, which amounts to assuming that all relevant $x \ll 1$ and selection is important only over many generations. Note the similarity of this equation to Eq. (1) governing the fitness distribution of the population. In analogy to Sec. I we can expand $\phi_z(x+\delta,t)$ inside the integral if the mutation rate is large and $K(\delta)$ falls of rapidly with $|\delta|$ and obtain

$$\partial_t \phi_z(x,t) = x\phi_z(x,t) - \sigma^2\partial_x\phi_z(x,t) + \frac{\mu\langle\delta^2\rangle}{2}\partial_x^2\phi_z(x,t) - \phi_z^2(x,t) \tag{15}$$

where $\sigma^2 = v - \mu\langle\delta\rangle$ accounts for both the actual translation of the population along the fitness axis and the mutation load. Defining $D = \frac{\mu\langle\delta^2\rangle}{2}$, we have the following equation for the generating function

$$\partial_t \phi_z(x,t) = x\phi_z(x,t) - \sigma^2\partial_x\phi_z(x,t) + D\partial_x^2\phi_z(x,t) - \phi_z^2(x,t) \tag{16}$$

with initial condition $\phi_z(x,0) = 1 - e^{-z}$. This equation allows two approximate solutions for small and large x. For small x, $\phi_z(x,t)$ is small we can neglect the non-linearity. Conversely, at large x, we find that $\phi_z(x,t)$ is large and the dominant balance is between the terms $x\phi_z(x,t)$ and $\phi_z^2(x,t)$ while the first and second derivative can be neglected. Using these approximations and assuming $z \ll 1$, we can solve Eq. (16) for small x and large x and match the two asymptotic solutions.

$$\phi_z(x,t) \approx \begin{cases} ze^{xt + \frac{Dt^3}{3} - \frac{\sigma^2 t^2}{2}} & x < x^* \\ x & x > x^* \end{cases} \tag{17}$$

The crossover is approximately at $x^* = \frac{\sigma^2 t}{2} - \frac{Dt^2}{3} - \frac{\log z/x^*}{t}$. Note that the branching process description is only valid for a time before population size constraints have to be imposed and lineages cease to be approximately independent [1, 2]. At short times, however, it is a valid approximate description of the number of offspring of a marked individual. We are interested foremost in the distribution of offspring averaged over the population $c(x)$, see Eq. (5). For short times with $x_c < x^*$ we have

$$\Phi(z,t) = zAe^{-\frac{\sigma^2 t^2}{2} + \frac{Dt^3}{3}}\int^{x_c} dx \mathrm{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x}{D^{\frac{1}{3}}}\right)e^{x\left(t - \frac{\sigma^2}{2D}\right)} \tag{18}$$

We see that the behavior of this average changes qualitatively as $t > \frac{\sigma^2}{2D}$ when it starts to be boundary dominated. Assuming $x^* < x_c$, i.e., the saturation of $\phi_z(x,t)$ happens within the validity of the Airy function solution, we have

$$\Phi(z,t) \approx zAe^{-\frac{\sigma^2 t^2}{2} + \frac{Dt^3}{3}} \int^{x^*} dx \mathrm{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x}{D^{\frac{1}{3}}}\right) e^{x\left(t - \frac{\sigma^2}{2D}\right)} + x^* A \int_{x^*}^{x_c} dx \mathrm{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x}{D^{\frac{1}{3}}}\right) e^{-\frac{x\sigma^2}{2D}} \tag{19}$$

Both of these integrals are dominated by a narrow region in the vicinity of x^*. Plugging in the definition of x^*, we find

$$\Phi(z,t) \approx x^* Ae^{-\frac{x^*\sigma^2}{2D}} \left[\int^0 d\delta x \mathrm{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x^* + \delta x}{D^{\frac{1}{3}}}\right) e^{\delta x(t - \frac{\sigma^2}{2D})} + \int_0 d\delta x \mathrm{Ai}\left(\frac{\sigma^4}{4D^{\frac{4}{3}}} - \frac{x^* + \delta x}{D^{\frac{1}{3}}}\right) e^{-\frac{\delta x\sigma^2}{2D}}\right] \tag{20}$$

$$= C(t,x^*)e^{-\frac{x^*\sigma^2}{2D}} = C(t,x^*)e^{\frac{\sigma^2 \log z/x^*}{2Dt} + \frac{\sigma^4 t}{4D} - \frac{\sigma^2 t^2}{6}} = E(t,x^*)z^{\frac{\sigma^2}{2Dt}}$$

where $C(t,x^*)$ and $E(t,x^*)$ depend only weakly on x^*. The kth derivative of $\Phi(z,t)$ with respect to z is therefore

$$\partial_z^k \Phi(z,t) \approx z^{-k}\Phi(z,t)\prod_{i=1}^{k-1}\left(\frac{\sigma^2}{2Dt} - i\right) = (-1)^k z^{-k}\Phi(z,t)\prod_{i=0}^{k-1}\left(i - \frac{\sigma^2}{2Dt}\right) \tag{21}$$

We can now evaluate the probability Q_k that a set of k lineages merges into one:

$$Q_k(t) \approx -\frac{N}{\Gamma(k)}\int_0^\infty dz z^{k-1}e^{-N\Phi(z,t)}(-1)^k\partial_z^k\Phi(z,t)$$

$$\approx -\frac{\prod_{i=0}^{k-1}(i - \frac{\sigma^2}{2Dt})}{\Gamma(k)}\int_0^\infty dz z^{-1}e^{-N\Phi(z)}N\Phi(z,t) \tag{22}$$

$$\approx -\frac{2Dt}{\sigma^2}\frac{\prod_{i=0}^{k-1}(i - \frac{\sigma^2}{2Dt})}{\Gamma(k)}\int_0^\infty d\phi\, e^{-\phi} = \frac{\prod_{i=1}^{k-1}(i - \frac{\sigma^2}{2Dt})}{\Gamma(k)} \quad .$$

Defining τ as $\tau = \frac{2Dt}{\sigma^2}$, we have

$$Q_k(t) = \frac{\Gamma(k - \tau^{-1})}{\Gamma(1 - \tau^{-1})\Gamma(k)} \quad . \tag{23}$$

For large τ, all Q_k become equal, which simply means that the population has coalesced into its common ancestor. For $t = \frac{\sigma^2}{2D} + \delta t$, we can differentiate with respect to δt and obtain

$$q_k = \partial_t Q_k(t)|_{t=\sigma^2/2D} = \frac{2D}{\sigma^2}\frac{1}{k-1} \quad , \tag{24}$$

which are the Bolthausen-Sznitman merger rates.

Similarly, we can calculate more general merger probabilities such as the probability that k out of b lineages have merged, while all other $b - k$ lineages trace back to distinct ancestors. Making again the assumption of independent lineages and accounting for the possibilities of choosing $b - k + 1$ ordered lineages as ancestors, we have

$$\Lambda_{b,k} = \frac{\Gamma(N+1)}{\Gamma(N-b+k)}\left\langle\left(\sum_j n_j\right)^{-b} n_1^k \prod_{i=1}^{b-k} n_{i+1}\right\rangle$$

$$\approx \frac{\Gamma(N+1)}{\Gamma(N-b+k)\Gamma(b)}\int_0^\infty dz\, z^{b-1}\langle n^k e^{-zn}\rangle\langle n e^{-zn}\rangle^{b-k}\langle e^{-zn}\rangle^{N-b+k-1} \tag{25}$$

$$= \frac{\Gamma(N+1)}{\Gamma(N-b+k)\Gamma(b)}\int_0^\infty dz\, z^{b-1}(-1)^{k+1}[\partial_z^k\Phi(z,t)][\partial_z\Phi(z,t)]^{b-k}e^{-(N-b+k-1)\Phi}$$

Using $\partial_t\Phi(z,t) = \frac{\sigma^2}{2Dt}z^{-1}\Phi(z,t)$, defining $\gamma = \frac{\sigma^2}{2Dt}$, and assuming $N \gg b$, we have

$$\Lambda_{b,k} = \frac{\gamma^{b-k}\Gamma(N+1)}{\Gamma(N-b+k)\Gamma(b)}\int_0^\infty dz\, z^{k-1}(-1)^{k+1}[\partial_z^k\Phi(z,t)]\Phi(z,t)^{b-k}e^{-(N-b+k-1)\Phi}$$

$$\approx -\prod_{i=0}^{k-1}(i - \gamma)\frac{\gamma^{b-k}}{\Gamma(b)}\int_0^\infty dz\, z^{-1}[N\Phi(z,t)]^{b-k+1}e^{-N\Phi} \tag{26}$$

Changing integration variables to $\phi = N\Phi(z,t)$, we find

$$\Lambda_{b,k} \approx -\prod_{i=0}^{k-1}(i-\gamma)\frac{\gamma^{b-k-1}}{\Gamma(b)}\int_0^\infty d\phi\,\phi^{b-k}e^{-\phi}$$

$$\approx -\prod_{i=0}^{k-1}(i-\gamma)\frac{\gamma^{b-k-1}\Gamma(b-k+1)}{\Gamma(b)} = \gamma^{b-k}\frac{\Gamma(k-\gamma)(b-k)!}{\Gamma(1-\gamma)(b-1)!} \tag{27}$$

For $t = \frac{\sigma^2}{2D} + \delta t$ with $\delta t \ll \frac{\sigma^2}{2D}$, $\gamma \approx 1$ and we find

$$\Lambda_{b,k} \approx \delta t\frac{2D}{\sigma^2}\frac{(k-2)!(b-k)!}{(b-1)!} = \delta t\lambda_{b,k} \tag{28}$$

which are the general BSC rates $\lambda_{b,k}$.

We have assumed a steady distribution with a well-defined cutoff at x_c. The actual high fitness end of the population fluctuates quite a bit, which would result in large population size fluctuations if the population size was not tightly controlled via a mean fitness that is catching up. This feed back via the mean fitness, however, is delayed and does not affect the dynamics at the nose instantaneously. The mapping to the FKPP models discussed below is better suited to describe the effects of fluctuations of the nose on coalescence.

IV. RELATION TO FKPP MODELS

After the equilibration time T_{eq} the ancestors of all extant individuals are located in a region of width $\propto D^{1/3}$ in the nose of the fitness distribution roughly $x_c \approx \frac{\sigma^4}{4D}$ ahead of the mean. Within this region, the growth rate does not vary much and the deterministic population distribution is a solution of

$$D\frac{\partial c(x)}{\partial x^2} - \sigma^2\frac{\partial c(x)}{\partial x} + x_c c(x) = 0 \tag{29}$$

which has an exponential solution $c(x) \sim e^{-x\gamma}$ with $\sigma^2(\gamma) = \frac{x_c}{\gamma} + D\gamma$. Waves of this form select the minimal velocity corrected by a cut-off [8], which corresponds to $\sigma^2 = 2\sqrt{x_c D}$ and $\gamma_0 = \sigma^2/2D$, consistent with [2]. Brunet et al. [9, 10] have argued that the coalescence process of such waves is of Bolthausen-Sznitman type with a typical time scale given by

$$T_c = \frac{\log^3 n}{\pi^2\gamma_0^3\alpha} = \frac{\log^3 n}{2\pi^2 x_c} \tag{30}$$

where $\alpha = 2x_c\gamma_0^{-3}$ is the second derivative of $\sigma^2(\gamma)$ with respect to γ at $\gamma = \gamma_0$ and n is the size of the population relevant for the FKPP wave. In our case, only the part of the wave that has a substantial chance of being the common ancestor of the population is relevant and this part is located in an area of with $\mathcal{O}(D^{1/3})$ away from the tip. The number of individuals in this range is

$$\log n \propto \log\int_{-D^{1/3}}^0 dx\,e^{-\frac{x\sigma^2}{2D}} \propto \frac{\sigma^2}{D^{2/3}} \tag{31}$$

Substituting this into Eq. (30) and using $x_c = \sigma^4/4D$, we find

$$T_c \propto \frac{\sigma^2}{D} \tag{32}$$

in agreement with the above findings. Furthermore, this argument supports our finding of Bolthausen-Sznitman coalescence after the branching process description is no longer valid.

V. THE BSC IN TUNED MODELS

The calculation giving rise to Eq. 5 of the main text was done for a specific model of selection and mutation. We now want to argue that similar considerations hold for a large class of related models. The evolution and composition

of traveling evolutionary waves is governed by the fitness distribution in the population, $c(x)$, and the probability that an individual at fitness x will take over the entire population $u(x)$ [1]. The deterministic and linear parts (i.e., up a cutoff) of $u(x)$ and $c(x)$ are left and right eigenfunctions of the same evolution operator $\mathcal{L}$. The distribution $c(x)$ is a localized pulse, while $u(x)$ is an increasing function of x (high fitness individuals have a higher chance of spreading than low fitness individuals).

Looking forward in time, $u(x)$ is the probability that a lineage at x is the common ancestor of the entire population in the distant future. Before the descendants of a single lineage take over the population, they spread evenly across the fitness distribution. Once spread out, every individual is equally likely to be the common ancestor and the chance that one descendant of this lineage fixes is given by the fraction $f(x)$ of the total population made up by descendants of the lineage. Hence $u(x)$ is roughly the product of the probability that the lineage survives up to the time when it's descendants are equally distributed across the population (establishment), and the expected fraction $f(x)$ conditional on establishment. Since only large values of $x \gg \sigma$ are relevant and the establishment probability in tuned models is $p_{est}(x) \sim x/2$, we expect $f(x) \approx 2u(x)/x$.

To analyze the coalescent properties, we need to know the distribution $p(f)$ of population fractions that trace back to single individuals a time T_{eq} in the past, where T_{eq} is the equilibration time of individuals across the population ($T_{eq} \approx \sigma^2/2D$ in our case).

$$p(f) = c(x)p_{est}(x)\left(\frac{df}{dx}\right)^{-1} \tag{33}$$

If the evolution operator $\mathcal{L}$ is dominated by selection all individuals that give rise to large f come from the high-fitness tail. In this region, $f(x)$ rises from very low values to $\mathcal{O}(1)$, while the product of $f(x)c(x)$ (or equivalently $u(x)c(x)$) is roughly constant (viewed as a function of f) [1]. Furthermore, $f(x)$ grows exponentially with a power of x, implying that $f'(x) = g(x)f(x)$ where $\log g(x)$ is a slowly varying function of x. Using $f'(x) \sim f(x)$ and $c(x) \sim f(x)^{-1}$, we find

$$p(f) \sim f^{-2} \tag{34}$$

Given this effective off-spring distribution, we can calculate the coalescence probabilities

$$\lambda_{b,k} \sim \int df\, p(f)f^k(1-f)^{b-k} = \frac{(k-2!)(b-k)!}{(b-1)!} \tag{35}$$

which are proportional to the BSC rates. The constant of proportionality depends on the details of the stochastic dynamics which we have calculated above by other means (see also [1] for an alternative calculation of $Q_2(T_{eq}) = \int dx c(x)u^2(x)$).

Distribution of ancestors

If $u(x)$ is the probability that a lineage with log-fitness x survives to the present, the log-fitness values of common ancestors of the population are distributed as $c(x)u(x)$. Now consider the MRCA common ancestor which lived a long time in the past. By definition, the MRCA gives rise to at least two branches that survive to the present. We therefore expect that the log-fitness of an MRCA that gives rise to k surviving branches should be approximately distributed as $c(x)u^k(x)$.

We measured the distribution of log-fitness of all common ancestors, CA, of the population and determined $u(x) =$ CA$(x)/c(x)$, see Fig. 5A of the main text. We then compared the distribution of the log-fitness of the MRCAs of the population to $c(x)u^2(x)$, finding good agreement with simulations. The data is too noisy to distinguish different $c(x)u^k(x)$ for different $k > 2$.

VI. LAPLACE TRANSFORM OF PAIR COALESCENCE TIMES

Instead of simulating individuals, one can also track the population density spread out on a 2 dimensional grid. To characterize diversity in an adapting population, we designate one dimension as the log-fitness axis, while the dynamics along the other direction is neutral direction. This keeps a record of the pair-coalescence time distribution as follows.

The distribution of pairwise distance between individuals along the neutral axis can be decomposed into the distribution of pairwise coalescence times and the distribution of distances along the neutral axis, given the coalescence

time T_2. If individuals hop left or right with probability $\mu/2$ each, the distance distribution along the neutral direction follows a diffusion equation with diffusion constant $D = \mu/2$. Hence the averaged distribution over all pairwise differences along the neutral coordinate should be

$$P(\Delta x) = \int_0^\infty dT P(T) \frac{e^{-\frac{\Delta x^2}{2DT}}}{\sqrt{2\pi DT}} \tag{36}$$

Fourier transforming this distribution in Δx results in

$$\hat{P}(k) = \int_0^\infty dT P(T) e^{-\frac{k^2 DT}{2}} \tag{37}$$

Hence the Fourier transform of the pair-wise distance function is the Laplace transform of the pair-coalescence time in the variable $k^2 D/2$.

We know from the individual based simulations that the coalescence process is essentially broken into two phases, a waiting period during which nothing is happening, followed by an exponential decay with charactistic time T_c. The Laplace transform of such a function is

$$\hat{P}(z) = \frac{e^{-T_{delay} z}}{1 + T_c z} \tag{38}$$

These two parameters can be easily extracted by fitting this function to simulation results. Note that this Laplace transform with $z = \mu$ equals the probability of identity by state of two randomly chosen individuals.

VII. SIMULATIONS

A. Forward simulations of adapting populations

In our stochastic simulations, each individual of the population is characterized by a log-fitness that determines the expected number of offspring of this individual in the following generation. The number of offspring is Poisson distributed with a mean $\exp(y - \lambda)$, where $\lambda = \bar{y} - (1 - N/N_0)$ is a density regulating factor that keeps the population size approximately at N_0. This type of density regulation is indistinguishable from a Fisher-Wright model with exactly constant population size if N_0 is large, in which case the actual population size deviates from N_0 only by fluctuations of order $\sqrt{N_0} \ll N_0$. Individual fitness y is updated via mutations drawn from a specified distribution $K(\delta)$ (in most cases a Gaussian, but we use other distributions too). The population is propagated asexually.

As the population evolves, we keep track of each individual's parent and construct a genealogy that links each individual in the present generation to all its ancestors. All nodes of the genealogy that don't have any offspring in the present population are deleted. From the genealogy, properties like the pair coalescent times can be easily calculated by tracing the lineages of randomly chosen individuals until they meet. Similarly, the most recent common ancestor of a population sample or the entire population can be determined this way.

Nodes in the genealogy also store how many individuals descent from them. This is convenient when calculating the site frequency spectra since a mutation that happens in a particular node of the genealogy is going to be present in all its descendants (ignoring back mutations). Since neutral mutations don't affect the population dynamics or the genealogies, we can calculate the neutral SFS simply by summing over all opportunities for neutral mutations in the genealogy.

B. Bolthausen-Sznitman coalescent simulations

The direct Bolthausen-Sznitman coalescent simulations follow a backward process starting with a population of N individuals. The time of the next merger event is drawn from an exponential distribution with rate given by Eq. 8 in the main text. Then the type of event and the k individuals that merge are chosen according to Eq. 7 of the main text. This process is repeated until all lineages have coalesced.

C. Measuring the Laplace transform of $P(T_2)$ in large populations

To measure the Laplace transform of the pair coalescence time distribution, we represent the population as a cloud on a two dimensional grid. This cloud spreads by diffusion in all directions, but is selected along one direction of grid.

Diffusion is approximated as a discrete hopping process with different hopping rates in the selected and unselected directions. As explained above (Eq. (36)), the distribution of the cloud in the direction orthogonal to the direction of selection is a measure of the pair coalescence time. This distribution is measured and saved for further analysis.

D.   Code

The source code of programs and scripts is available from the author's website:

http://www.eb.tuebingen.mpg.de/research/research-groups/richard-neher/theoretical-population-genetics.html

---

[1] Hallatschek, O. (2011) The noisy edge of traveling waves. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1783–7.

[2] Cohen, E, Kessler, D. A, & Levine, H. (2005) Front propagation up a reaction rate gradient. *Phys Rev E Stat Nonlin Soft Matter Phys* **72**, 066126.

[3] Rouzine, I. M, Wakeley, J, & Coffin, J. M. (2003) The solitary wave of asexual evolution. *Proc Natl Acad Sci USA* **100**, 587–92.

[4] Desai, M. M & Fisher, D. S. (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–98.

[5] Desai, M. M, Walczak, A. M, & Fisher, D. S. (2012) Genetic diversity and the structure of genealogies in rapidly adapting populations. *arXiv:1208.3381*.

[6] Berestycki, N. (2009) Recent progress in coalescent theory. *arXiv* **math.PR/0909.3985**.

[7] Yule, G. U. (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos Trans R Soc Lond, B, Biol Sci* **213**, 21–87.

[8] Brunet, E & Derrida, B. (1997) Shift in the velocity of a front due to a cutoff. *Physical Review E*.

[9] Brunet, E, Derrida, B, Mueller, A. H, & Munier, S. (2006) Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Physical review E, Statistical, nonlinear, and soft matter physics* **73**, 056126.

[10] Brunet, E, Derrida, B, Mueller, A. H, & Munier, S. (2007) Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Physical review E, Statistical, nonlinear, and soft matter physics* **76**, 041104.

[11] Hudson, R. R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–8.
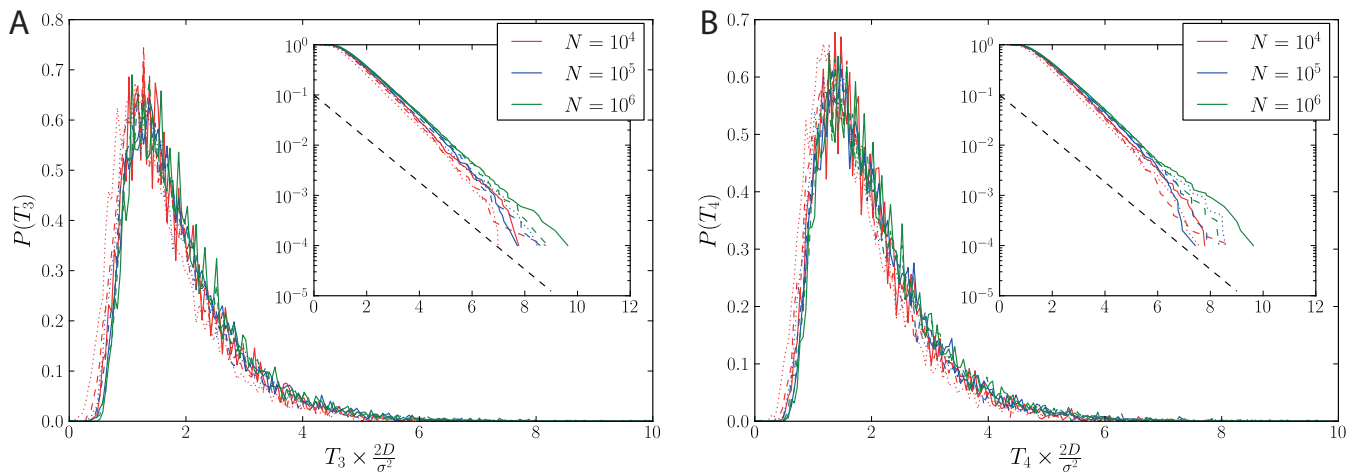
Supplementary Figure S1: The distribution of the time to the most recent common ancestor of three individuals (panel A), and four individuals (panel B) in a model where mutational effects are normally distributed and mutations are frequent; compare Fig. 3 of the main text. Different line styles correspond to $s = 0.01$ (solid), $s = 0.001$ (dashed), and $s = 0.0001$ (dotted), while the mutation rate is $\mu = 1$. For each parameter combination, random pairs are sampled at 10000 time points $2s^{-2/3}$ generations apart
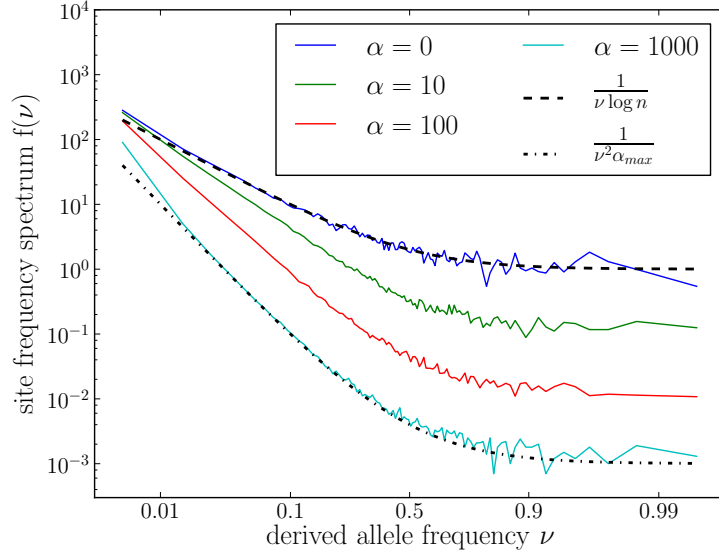
Supplementary Figure S2: The site frequency spectrum of neutral derived mutations in Kingman's coalescent with constant ($\alpha = 0$) and exponentially growing population sizes ($\alpha > 0$ is the growth rate measured in units of N^{-1}). The genealogies are produced with the program `ms` [11]. The black lines show the theoretical expectation for a population of constant size and a rapidly expanding population. The x-axis is scaled as in Fig. 4 of the main text.
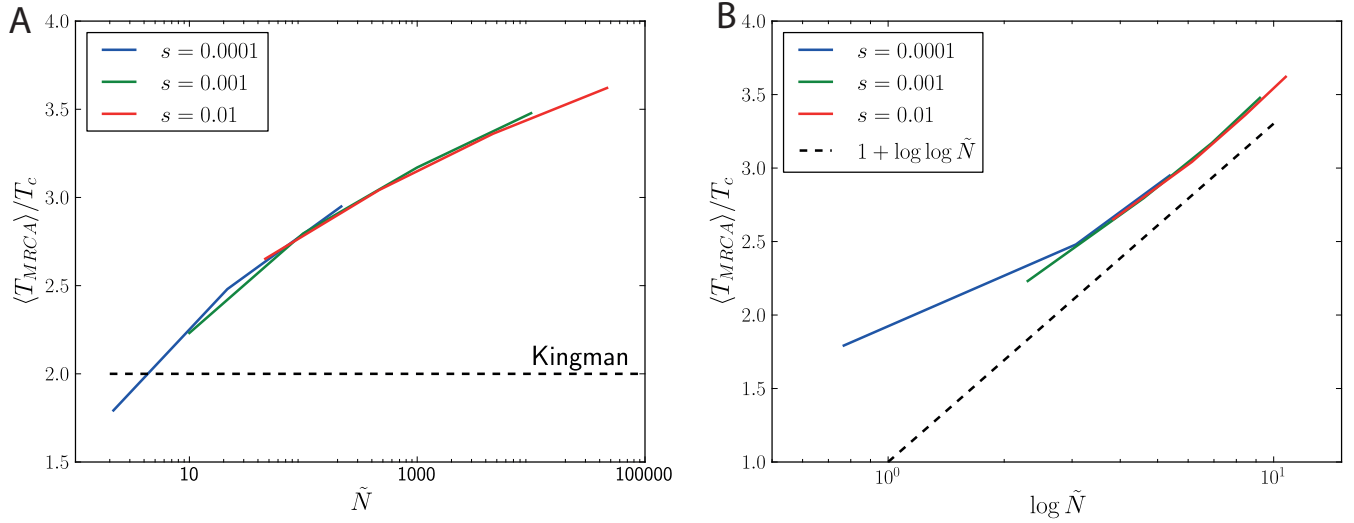

Supplementary Figure S3: The average time to the most recent common ancestor of the whole population increases relative to the time scale of coalescence with the population size $\tilde{N} = ND^{\frac{1}{3}}$. This is a well known feature of the Bolthausen-Sznitman coalescent where one expects $\langle T_{MRCA} \rangle = T_c \log \log N$. Panel B compares $\langle T_{MRCA} \rangle / T_c$ to $1 + \log \log \tilde{N}$, where the additional 1 is necessary to account for $T_{delay} \approx T_c$ before coalescence sets in.
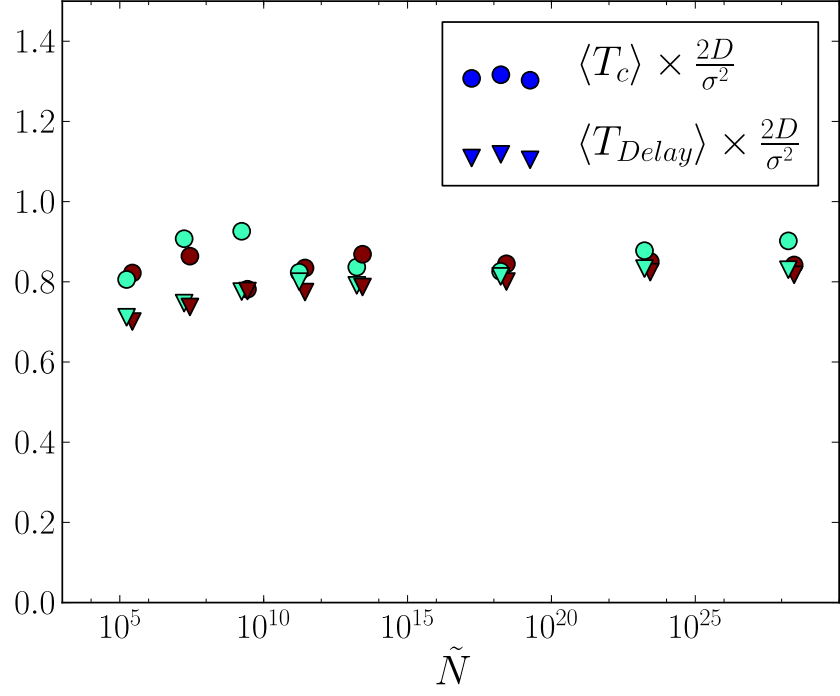
Supplementary Figure S4: The delay T_{delay} and the coalescence time scale T_c can be extracted from the Laplace transform of $P(T_2)$ for very large populations and confirms that $T_c \approx \frac{\sigma^2}{2D}$.
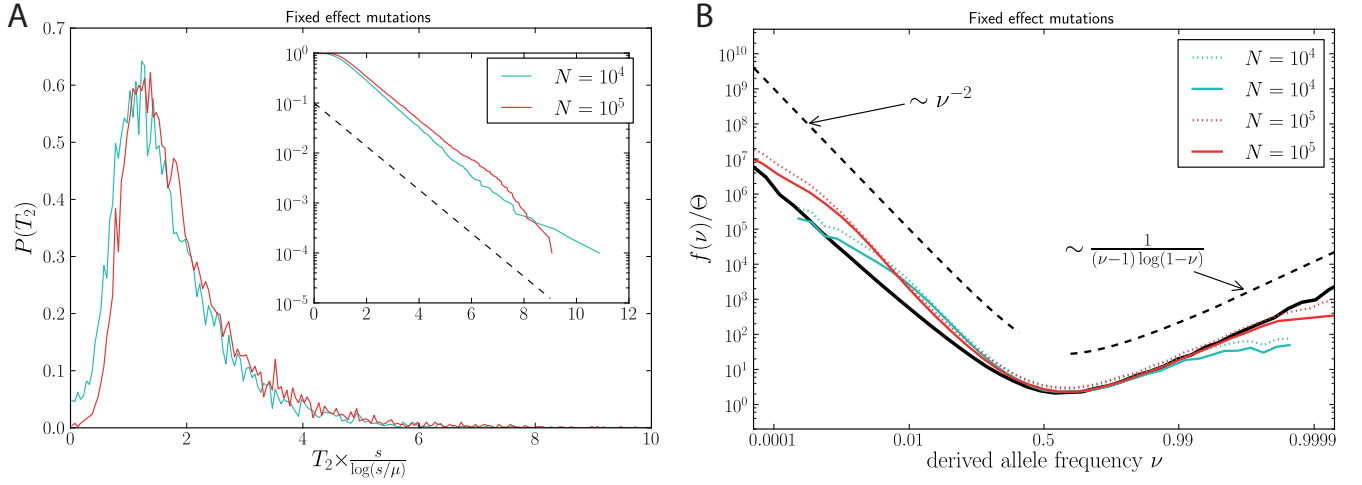
Supplementary Figure S5: Results for an alternative models of adaptation, in which all mutations confer the same selective advantage s and the mutation rate satisfies $\mu \ll s$. Panel A shows the distribution of pair coalescent times, panel B shows the site frequency spectra and the comparison to the Bolthausen-Sznitman coalescent. The times are rescaled with the prediction for the coalescent time $\langle T_2 \rangle \approx s^{-1} \log(s\mu^{-1})$ by Desai et al. [5]. Solid lines correspond to $\mu/s = 0.01$, dotted lines to $\mu/s = 0.1$, while $s = 0.01$.
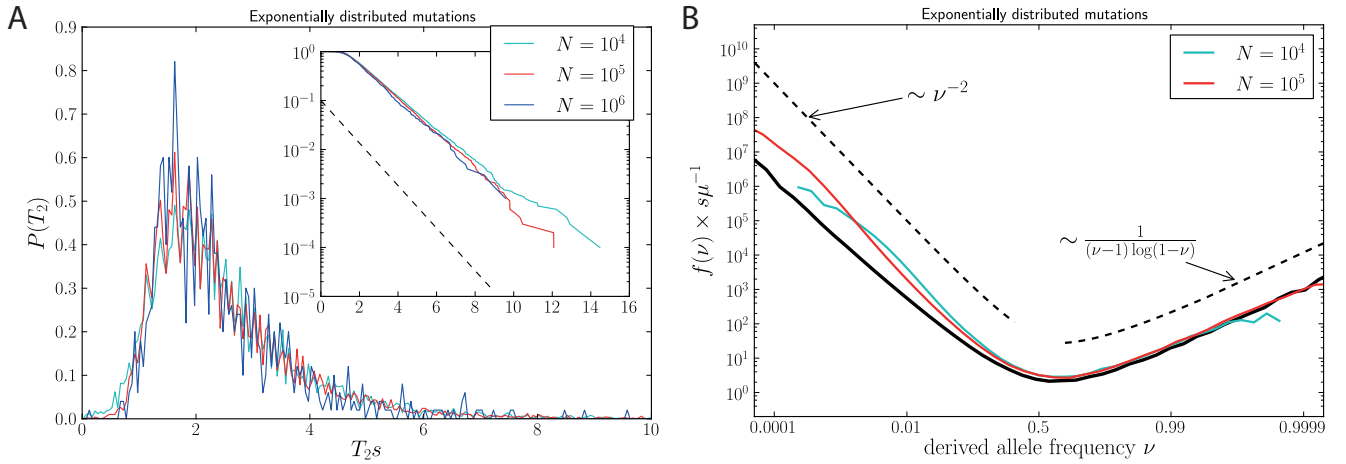
Supplementary Figure S6: Results for an alternative model of adaptation, in which the selective advantage of new mutations is drawn from an exponential distribution with mean s, while the mutation rate satisfies $\mu \ll s$. Panel A shows the distribution of pair coalescent times, panel B shows the site frequency spectra and the comparison to the Bolthausen-Sznitman. In the absence of a prediction for the dependence of T_2 on Ns or μs^{-1}, we rescale times by $s = 0.01$. The mutation rate equals $\mu = 0.1s$.