Horizontal gene transfer and homologous recombination in rapidly adapting populations

James T. Ferrare[1], Daniel P. Wong[2**] Benjamin H. Good[1,2,3,4,†]**

[1] *Biophysics Program, Stanford University, Stanford, CA 94305, USA*
[2] *Department of Applied Physics, Stanford University, Stanford, CA 94305, USA*
[3] *Department of Biology, Stanford University, Stanford, CA 94305, USA*
[4] *Chan Zuckerberg Biohub – San Francisco, San Francisco, CA 94158, USA*

[**] *These authors contributed equally to this work*
[†] *Correspondence should be addressed to: B.H.G. (bhgood@stanford.edu)*

Abstract

Horizontal gene transfer is a ubiquitous force in microbial evolution. Although HGT is commonly associated with the transfer of whole genes and operons, it can also play a more subtle role reshuffling genetic variants between closely related strains. This process of recombination is especially important in adapting populations that harbor many competing genetic variants. Here, we develop a mathematical framework to study the interplay between localized recombination and rapid adaptation in large populations. We find that recombination can become the dominant force driving the success of genetic variants even when it is extremely rare relative to mutation. We show that this crossover depends on the population size and distribution of transfer lengths and is often driven by the repeated transfer of genetically colocalized mutations. These successful variants can leave unique signatures of recombination in longitudinal data and strongly bias inferred recombination rates and length scales. They also drive large fluctuations in fitness and genetic diversity as they become common while remaining difficult to detect from summary statistics like linkage disequilibrium. Collectively, our theory demonstrates that recombination introduces novel non-equilibrium behavior that cannot be captured in existing frameworks.

Main Text

The horizontal exchange of genetic material—commonly referred to as horizontal gene transfer (HGT)— is an important driver of adaptive evolution in the microbial world (1). While HGT is often associated with the transfer of novel genes and pathways (2, 3), it can also facilitate the exchange of homologous segments of DNA between pairs of closely related strains (4, 5). These segments can be incorporated into a recipient cell's genome via homologous recombination, transferring new mutations and overwriting existing genetic variation. Within a competing population, this process acts to reshuffle genetic variants, representing a minimal form of the genome-wide exchange that occurs during meiotic recombination. As a result, HGT can accelerate adaptive evolution by rescuing beneficial variants from poor genetic backgrounds and bringing together beneficial combinations of mutations (6, 7). Examples are common in microbial systems, including the proliferation of antibiotic resistance (8–11), evolution of novel metabolic pathways (12, 13), and adaptation to environmental stress (14). But despite the ubiquity of this localized recombination across microbial species, its effect on the long-term success of genetic variants remains poorly understood.

Classical theory predicts that recombination will play an important role shaping genetic diversity in

adapting populations (15–17). This work applies when beneficial mutations are rare and do not strongly interfere with each other. In these scenarios, a sweeping beneficial variant can recombine onto multiple genetic backgrounds, allowing any linked variants to hitchhike to larger frequencies. Yet, microbes often reside in large populations where many beneficial mutations can arise and compete on separate genetic backgrounds, preventing any one from quickly taking over (18–23). For instance, in the human gut, the typical bacterial population size and per genome mutation rate are $N \sim 10^{12}$ (24, 25) and $U \sim 10^{-3}$ (26), ensuring enormous amounts of genetic variation are produced each generation. Even if beneficial mutations occur at a modest rate $U_b \sim 10^{-6}$, corresponding to 1 beneficial site per 1000 in a genome $f_b \sim 10^{-3}$ (21), then $NU_b \sim 10^6$ beneficial mutations are produced each day ($\sim$ 1 generation/day (25)). In the absence of recombination, most of these adaptive mutations are wasted on poor genetic backgrounds that cannot take over in the long-term. By unlinking beneficial mutations from their original genetic backgrounds, recombination can in principle alleviate this process of clonal interference (27). Yet, the interplay between clonal interference and recombination is not well characterized in many empirical settings of interest.

Recent work has considered the role of localized recombination in rapidly adapting populations ($NU_b \gg 1$), but focused on scenarios in which the transfer rate of a beneficial locus is larger than the typical beneficial effect size (7, 28). In this regime, any new adaptive allele that arises, and reaches an appreciable frequency, becomes immediately unlinked from its original genetic background. As a consequence, in this limit there is no genome-wide clonal structure in a population. Yet, longitudinal sequencing of many microbial populations, including those in the gut, clearly evince mostly clonal structure and strong linkage at detectable frequencies (29, 30). Moreover, recombination is often observed to occur at low relative rates —often comparable with those of mutation —where linkage is expected to play an important role. For example, at empirically observed recombination rates $r/\mu \sim 1$ and transfer lengths $\ell \sim 10^3$ in the human gut, a microbial genome recombines only once per $r \cdot L \sim 10^4$ generations while exchanging only a fraction of its genome $\overline{\ell}/L \sim 10^{-3}$ (31). This qualitative gap between our theoretical understanding and empirical observations leaves basic questions unresolved: how much localized recombination is necessary to have an appreciable effect on the tempo and mode of adaptive evolution in clonally interfering populations? What departures from clonality in genomic sequencing data does recombination leave in such populations? And how do answers to these questions depend on extrinsic factors like the population size, mutation and recombination rates, and selective effects?

A simple model of horizontal gene transfer and adaptation

To address these questions, we study rapidly adapting populations with weak to intermediate rates of localized recombination, where selection is comparatively strong enough to maintain clonal structure in the population. We consider a simple model of a population of N individuals that can acquire mutations at a total of L linked genetic loci. We initially focus on the most basic scenario where a large number of beneficial sites comprising a small fraction of total genomic loci $L \gg f_b \cdot L \gg 1$ are randomly distributed along the genome. We assume that epistasis is sufficiently weak so that the genetic positions of these beneficial loci as well as their mutation rate and fitness benefit remain approximately constant over the relevant timescales (we later relax these assumptions below).

Recombination events occur at a total per genome rate $r \cdot L$ when a random donor individual transfers a genetic segment of length ℓ to a random acceptor individual, which then integrates the segment into its genome, overwriting any existing variants. The probability that any focal site is transferred will be proportional to the length of this transferred segment ℓ/L which is drawn from the distribution of

transfer lengths $\lambda(\ell)$. This implies that the total transfer rate of a focal site is simply $r \cdot \overline{\ell}$, where $\overline{\ell} \equiv \int \ell \cdot \lambda(\ell) d\ell$ is the mean length of a transferred segment. Consistent with the minimal nature of HGT (31), we assume that the mean transfer length remains much smaller than the total genome length $\overline{\ell} \ll L$.

The success of a recombination event will depend on the fitness benefits of the mutations that are transferred. To begin, we consider the simplest possible distribution of fitness effects (DFE) where all mutations have the same fitness benefit s_b and occur at the same total rate U_b (we later relax this assumption as well). This allows us to assume that the fitness benefits of mutations at a large frequency and most likely to be transferred are the same as those produced by mutation. It also enables us to focus on a broad parameter regime motivated by empirically relevant scenarios in which clonal interference is pervasive and selection is relatively strong: $s_b \gg U_b \gg 1/N$. In the limit where the transfer rate becomes much larger than the typical benefit $r \cdot \overline{\ell} \gg s_b$, mutations at a large frequency will become effectively unlinked (7). In the opposite limit where recombination is extremely rare $r \cdot \overline{\ell} \ll 1/N$, this model reverts to classic models of asexual rapid adaptation with a stable distribution of fitness effects (32–34). This provides a natural null model to measure the effects of recombination. As we allow for HGT, a natural question emerges: how will recombination influence the fate of a new mutation?

We find that recombination can become the dominant force driving the success of genetic variants, even when it is extremely rare relative to mutation. We identify this crossover by measuring the fraction of successful mutations that take over on their original genetic backgrounds (Fig. 1A). In the asexual limit, the typical successful beneficial mutation must produce multiple additional linked mutations to outcompete other mutations and fix. As a result, most successful mutations will remain linked to their original genetic background and the rate of fixation θ_r will remain near its asexual value θ_a. As recombination becomes the dominant force, most successful mutations will take over on a background different from their original one and θ_r will become appreciably larger than θ_a. Naively, we might expect this transition to occur when a mutation's chance of producing and acquiring a linked mutation balance $r \cdot \overline{\ell} \sim U_b$. This boundary also often coincides with another null expectation, where the per-base-pair recombination and mutation rates balance $r \sim \mu$, reflecting their equal nominal contributions to evolutionary dynamics.

Strikingly, in contrast to these expectations, we find a sharp transition to recombination dominated dynamics at extremely small recombination rates for a range of empirically motivated parameters (Fig. 1B). In particular, for the large population sizes relevant to the human gut microbiome, recombination can become the dominant force driving the success of genetic variants at rates as small as $r/\mu \lesssim 10^{-3}$ or much smaller depending on the mutation rate and distribution of transfer lengths. Interestingly, this crossover strongly depends on the population size, with the crossover value of r/μ decreasing roughly inversely with the population size N (Fig. 1B). This highlights how competition between linked genetic variants in large populations can dramatically amplify the role of recombination.

Our results also show that the role of recombination will scale with the typical fraction of the genome involved in a transfer event (Fig. 1C). In particular, the fixation rate of recombined mutations scales linearly with the typical transfer and genome lengths. As a consequence, order-of-magnitude differences in these values can shift the population towards recombination dominated dynamics. Interestingly, these changes can occur while the the overall transfer rate $r \cdot \overline{\ell}$ remains constant. This demonstrates that linkage within localized segments of the genome often remains important in recombining populations. These results also suggest that the strength of selection on individual

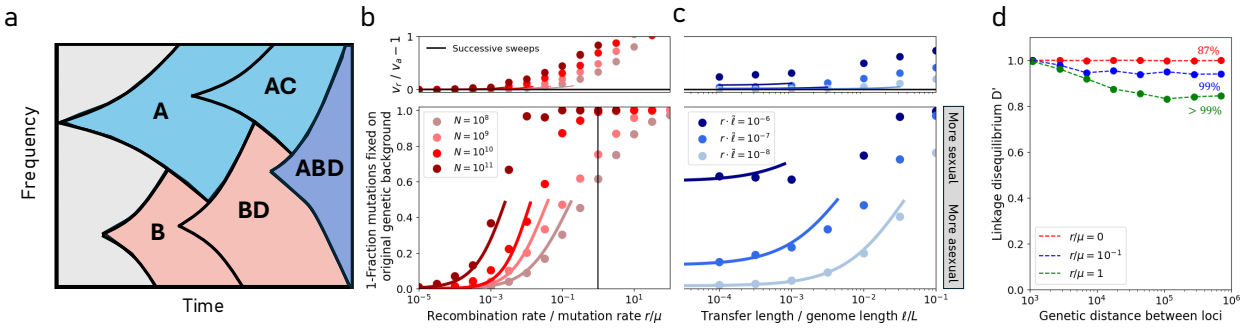


Figure 1: **Interference between linked mutations enhances the role of recombination. (a)** Mueller diagram with clonal interference and horizontal gene transfer. Mutations in separate clonal lineages (red and blue) compete and rise to high frequency on their original genetic background. A transfer event recombines mutation 'A' onto the 'BD' genotype, which takes over on this new background (purple). **(b,c)** The fraction of mutations that fix on a background different from their original one (bottom). The rate of adaptation of a sexual population vs. an asexual population with the same population size and distribution of fitness effects (top). Symbols denote the results of forward time simulations for $s_b = 5 \cdot 10^{-2}$, $\mu = 10^{-10}$, $f_b = 10^{-2}$, $\bar{\ell} = 10^2 - 10^5$, $L = 10^6$, $N = 10^8 - 10^{11}$ and $r = 10^{-8} - 10^{-13}$. Solid lines denote our theoretical predictions. **(d)** Normalized linkage disequilibrium versus genetic distance. Symbols denote average values from forward time simulations with the same parameters in panel b). Percentages denote the corresponding fraction fixed on a background different from their original one from panel b). Dashed lined are shown to aid visualization.

transfer events will scale with the length of the transferred segment, and thus will vary across the distribution of transfer lengths (see Discussion).

Finally, while HGT can strongly influence mutations' evolutionary trajectories, we find that the transition to recombination dominated dynamics can be difficult to detect from summary statistics like the rate of adaptation or linkage disequilibrium (Fig. 1D). Even while most mutations take over on a background different from their original one, the rate of adaptation increases only modestly from its asexual value [e.g. $v_r/v_a \lesssim 2$]. Moreover, genome-wide linkage is largely preserved, reflecting ongoing clonal evolutionary dynamics. These results highlight how recombination can play a dramatic role in the fate of individual genetic variants, while only subtly influencing the tempo and mode of adaptive evolution.

Theoretical model

To understand these results, we must follow the fate of a new mutation as it rises to large frequency and competes for dominance in the population. The fate of a new mutation will strongly depend on the fitness of its original genetic background. In the simplest scenario, recombination will act too slowly to break apart the mutation from this background before it either takes over or is purged from the population. Consequently, the typical successful mutation will need to produce multiple additional linked mutations to outcompete other mutations and fix. This implies that the distribution of fitness will largely be shaped by asexual competition. In these cases, previous work (34) has shown that it will approach a steady-state profile that increases in fitness at a roughly constant rate v_a (Fig. 2B). This distribution will have a characteristic width, x_{ca}, which also depends on the distribution of fitness effects and population size, and roughly coincides with the location of the fittest individuals in the population.

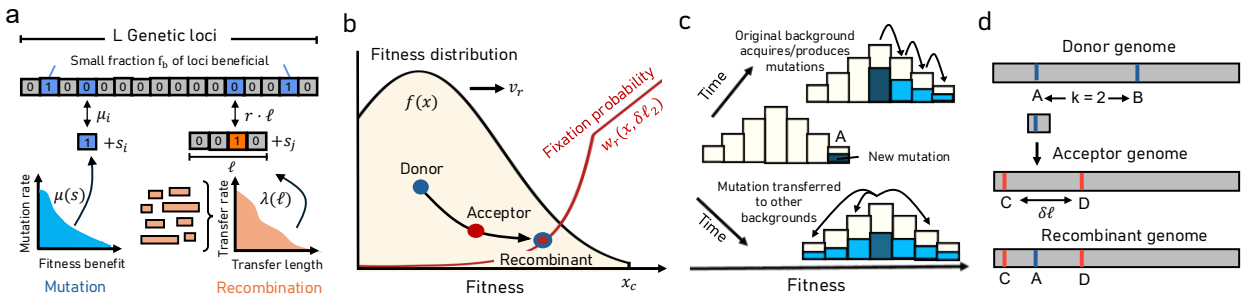


Figure 2: **Modeling horizontal gene transfer in rapidly adapting populations.** **(a)** Each genome has a fraction f_b of beneficial loci randomly distributed along L total linked biallelic loci. Each site i mutates at rate μ_i and confers a fitness advantage s_i. The rates and benefits of individual mutations are drawn from the distribution of fitness effects $\mu(s) \equiv \sum_{i \in L} \mu_i \cdot \delta(s - s_i)$. Mutations can also be transferred via HGT events, which occur at an overall rate $r \cdot L$. The total transfer rate of a polymorphic site scales as $r \int \ell \cdot \lambda(\ell)\, d\ell$, where $\lambda(\ell)$ is the distribution of transfer lengths. **(b)** Schematic of the asexual fitness distribution. The distribution takes a steady state shape with a typical maximum fitness x_c and advances at a roughly steady state rate v_r. Acceptor and donor individuals are sampled from the fitness distribution producing a recombinant offspring. The fixation of this offspring depends on its relative fitness x. **(c)** A new mutation 'A' arises in the nose of the fitness distribution. This mutation can take over on its original genetic background by producing or acquiring additional linked mutations (top). Alternatively, the mutation can take over by being transferred to another more fit genetic background (bottom). **(d)** Schematic of a horizontal gene transfer event. The donor genotype transfers one of its k unique mutations to the acceptor founding a new recombined genotype. Mutations on the acceptor genotype are separated by a distance $\delta\ell$.

A new mutation will arise on a genetic background with relative fitness drawn from $f(x)$ and compete with other genotypes to take over (Fig. 2C). Provided the mutation takes over on this background by producing linked mutations, the outcome of this competition can be described by the asexual conditional fixation probability $w_a(x)$. Previous work (34) has shown that the asexual conditional fixation probability satisfies the branching process recursion,

$$0 \approx \underbrace{x \cdot w_a(x)}_{\text{selection}} - \underbrace{v_a \cdot \partial_x w_a(x)}_{\text{adaptation of population}} - \underbrace{\frac{1}{2} \cdot w_a(x)^2}_{\text{genetic drift}} + \underbrace{U_b \cdot [w_a(x + s_b) - w_a(x)]}_{\text{mutation produced}}, \tag{1}$$

which represents a balance between the (i) the growth of a clone due to selection, (ii) the adaptation of the background population, (iii) the stochastic effects of genetic drift, and (iv) the production of further mutations. The overall substitution rate of mutations on their original background can then be obtained by averaging over the distribution of genetic backgrounds,

$$\underbrace{\theta_a}_{\text{asexual rate of fixation}} = \underbrace{NU_b}_{\text{mutation rate}} \cdot \int \underbrace{f(x) \cdot w_a(x + s_b)\, dx}_{\text{prob. background} \times \text{prob. fixation}}. \tag{2}$$

By extending Eq. (1) to include contributions from horizontal gene transfer, we can ask: when do recombined mutations start to fix on a par with those on their original background?

Recombination can influence the fate of a new mutation in several key ways. First, HGT events can transfer mutations from other genetic backgrounds to the new mutation's background. This process is

similar in effect to producing linked mutations, since the focal mutation can hitchhike to fixation with the linked mutations it acquires. The rate of these acquisition events will depend on the genetic diversity and descent of genotypes within the population. This is because an acceptor genotype cannot gain an advantage by acquiring a mutation that it already possesses. As a result, the overall rate at which the focal mutation's background acquires mutations will depend on the number of mutations unique to a random donor drawn from $f(x)$ (Fig. 2D). As we will show, successful transfers will generally involve an acceptor with greater fitness than a donor. In this case, provided recombination is sufficiently rare, previous work has shown that this distribution will be independent of the donor and acceptor's fitnesses and have an average value $\tilde{k}$ equal to the typical number of mutations separating the fittest individuals from those at the mean fitness x_c/s_b (35). For the empirically relevant scenarios we consider, this number will not be too large $1 \lesssim \tilde{k} \lesssim 10$, so that the typical transfer event involves at most a single mutation $\tilde{k} \ll L/\tilde{\ell}$.

However, even while acquisition events involving multiple mutations are rare, they can provide a multiplicative fitness benefit to the acceptor genome when they do occur. The rate of acquiring multiple mutations will depend on the genetic distances $\delta\ell$ between adjacent mutations on a random donor genotype (Fig. 2D). This distribution follows straightforwardly from the random uniform distribution of beneficial loci, $\Pr(\delta\ell) \propto e^{-k\cdot\delta\ell/L}$. Most transfer lengths will be smaller than this typical spacing, in which case $\Pr(\delta\ell)$ simplifies to a uniform distribution, allowing us to approximate the rate of acquiring n mutations from a random donor as $\tilde{R}_{n\leftarrow} \approx r \cdot L \cdot \int \lambda(\ell) \cdot \left(\frac{\tilde{k}\cdot\ell}{L}\right)^n d\ell$, where the left arrow denotes mutations incoming to the focal mutation's genome. This implies the role of recombination will depend on the balance between the rarity at which mutations are colocalized $\sim \int \lambda(\ell) \cdot (\ell/L)^n$ and the benefit they provide $\sim n \cdot s_b$.

The focal mutation can also take over by jumping to an entirely different genetic background itself. The contributions from this process will strongly depend on the distribution of acceptor backgrounds which will also be randomly drawn from $f(x)$ (Fig. 2C; bottom). Although most transfers will land the mutation on a less fit or roughly equivalent genetic background, rare transfers can lead to atypically fit backgrounds. The contributions from this process will depend on the genetic position of the focal mutation relative to other mutation's on its original background. For example, if a new mutation arises near another preexisting mutation, the two mutations' fate can become linked if they are likely to be transferred together. Since the dynamics are predominantly asexual, the genetic distance separating a new mutation from its nearest neighbor $\delta\ell_2$ will also be randomly sampled from $\Pr(\delta\ell)$. This ensures that for typical values of $\delta\ell_2$, the focal mutation will be overwhelmingly likely to be transferred alone at a total rate, $\tilde{R}_{1\rightarrow} = r \cdot \int_0^{\delta\ell_2} \ell \cdot \lambda(\ell)\, d\ell + r \cdot \int_{\delta\ell_2}^{L} \delta\ell_2 \cdot \lambda(\ell)\, d\ell$, where the right arrow denotes the focal mutation being transferred away from its present genotype. More rarely, the focal mutation will arise near a preexisting mutation, in which case the colocalized pair can be transferred together at a total rate, $\tilde{R}_{2\rightarrow} = r \cdot \int_{\delta\ell_2}^{L} \lambda(\ell) \cdot (\ell - \delta\ell_2)\, d\ell$. While colocalization events are rare, they can also provide a multiplicative fitness advantage to the acceptor genome. From the perspective of a new mutation, this represents an important difference from acquiring multiple mutations: the probability of becoming colocalized scales only linearly with the transferred length $\int \lambda(\ell) \cdot \ell/L$, after which the mutation can recoup a $2s_b$ benefit with each subsequent transfer. This requires us to extend the conditional fixation probability of a focal mutation to depend on the genetic distance to its nearest neighbor $w_r(x, \delta\ell_2)$. This conditional fixation probability will satisfy a generalized version of the branching process

recursion in Eq. (1),

$$0 \approx \underbrace{x \cdot w_r(x,\delta\ell_2)}_{\text{selection}} - \underbrace{v_a \cdot \partial_x w_r(x,\delta\ell_2)}_{\text{adaptation of the population}} \underbrace{- \frac{1}{2} \cdot w_r(x,\delta\ell_2)^2}_{\text{genetic drift}} + \underbrace{U_b \int_{\delta\ell_2}^{L} \Pr(\delta\ell) \cdot [w_r(x+s_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{distant mutation produced}}$$
$$+ \underbrace{U_b \int_{0}^{\delta\ell_2} \Pr(\delta\ell) \cdot [w_r(x+s_b,\delta\ell) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{more colocalized mutation produced}} + \underbrace{\sum_n \frac{\tilde{R}_{n\leftarrow}}{2} \int_{\delta l_2}^{L} \Pr(\delta\ell) \cdot [w_r(x+ns_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{distant mutation(s) acquired}}$$
$$+ \underbrace{\frac{\tilde{R}_{1\leftarrow}}{2} \int_0^{\ell} \frac{dz}{\ell} \int_z^{\delta l_2} \Pr(\delta\ell) \cdot [w_r(x+ns_b,\delta\ell) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{more colocalized mutation acquired}} + \underbrace{\sum_n \frac{\tilde{R}_{n\leftarrow}}{2} [w_r(x-ns_b,\delta\ell_2) - w_r(x,\delta\ell_2)]}_{\text{mutation(s) overwritten}}$$
$$+ \underbrace{\frac{\tilde{R}_{1\rightarrow}}{2} \int_0^{\ell} \frac{dz}{\ell} \int_z \Pr(\delta\ell) \int f(x) \cdot [w_r(x+s_b,\delta\ell) - w_r(x,\delta\ell)]\, dx}_{\text{focal mutation transferred alone}} + \underbrace{\frac{\tilde{R}_{2\rightarrow}}{2} \int f(x)\, [w_r(x+2s_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, dx}_{\text{focal mutation transferred as pair}}, \tag{3}$$

which now includes contributions from (v) the acquisition of one or more mutations, (vi) the overwriting of existing mutations via HGT, (vii) the transfer of the focal mutation alone to a new background, and (viii) the transfer of the focal mutation and a colocalized mutation to a new background. We have also neglected the possibility that the focal mutation becomes part of a colocalized triplet, which as we will discuss, can become relevant for extremely large population sizes.

The total substitution rate including contributions from recombination can then be obtained by averaging over the nearest neighbor distances and initial fitness backgrounds they could arise on,

$$\underbrace{\theta_r}_{\text{sexual rate of fixation}} = \underbrace{NU_b}_{\text{mutation rate}} \cdot \int \underbrace{\Pr(\delta\ell_2)\, d\delta\ell_2}_{\text{prob. distance to nearest mutation}} \int \underbrace{f(x) \cdot w_r(x+s_b,\delta\ell_2)\, dx}_{\text{prob. background} \times \text{ prob. fixation}}. \tag{4}$$

This provides a natural scale to measure the effect of recombination. When θ_r becomes much larger than θ_a, most mutations will take over on a background different from their original one (Fig. 1A). When this occurs, recombination will transition to become the dominant force driving mutations to fix and our assumptions about the shape of the distribution of fitness and genetic diversity will break down.

Theoretical results for a simple model of the distribution of transfer lengths

To understand this crossover, we start by considering a simple model of the distribution of transfer lengths, where all transferred segments have the same length $\tilde{\ell}$. This implies that the rate of acquiring n mutations will simply scale as $(\tilde{\ell}/L)^n$. It also allows us to clearly define colocalized pairs as mutations with a nearest neighbor distance smaller than the transfer length $\delta\ell_2 \lesssim \tilde{\ell}$ since both mutations will typically be transferred together at a total rate $\tilde{R}_{2\rightarrow} \approx r \cdot (\tilde{\ell} - \delta\ell_2)$. On the other hand, when the initial separation exceeds the transfer length $\delta\ell_2 \gtrsim \tilde{\ell}$, the new mutation will always be transferred alone at a total rate $\tilde{R}_{1\rightarrow} \approx r \cdot \tilde{\ell}$.

For small transfer lengths and population sizes $(NU_b)^{3/4} \lesssim L/\tilde{\ell}$ or overall numbers of beneficial sites $L \cdot f_b \lesssim \tilde{\ell}$, we find that the relative rate of fixation of recombined mutations initially scales as,

$$\log\left(\frac{\theta_r}{\theta_a}\right) \sim \frac{r \cdot \tilde{\ell} \cdot \tilde{k}}{U_b}. \tag{5}$$

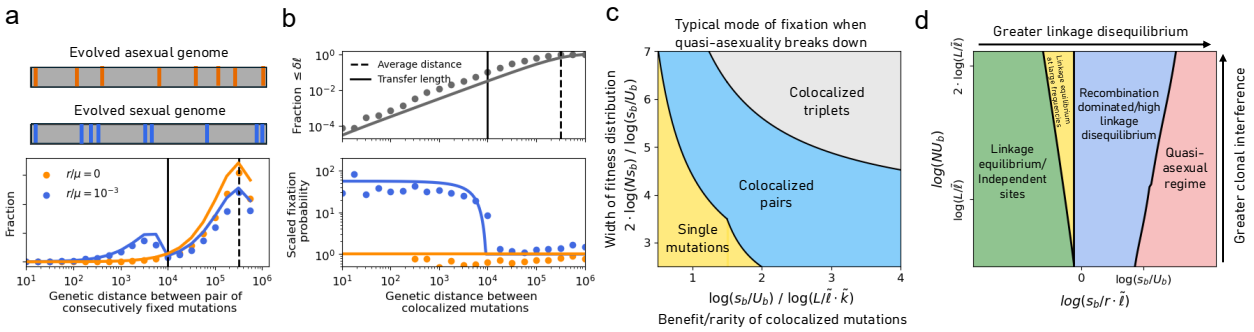


Figure 3: **Success of recombinants driven by rare colocalized transfers. (a)** Evolved asexual and sexual genomes (top). Distribution of genetic distances separating mutations fixed consecutively in time (bottom). Lines denote the results of simulations where $N = 10^{11}$, $s_b = 5 \cdot 10^{-2}, \tilde{\ell} = 10^4, L = 10^6, r = 10^{-12}, \mu = 10^{10}$, $f_b = 10^{-3}$. **(b)** Cumulative distribution function of the distance separating a random pair of beneficial loci (top). The probability that a random pair of mutations is colocalized within a transfer length is $\sim 10^{-2}$. Theoretical prediction for the relative fixation probability of a colocalized pair as a function of the genetic distance separating the pair of mutations (bottom). **(c)** Phase diagram showing the typical number of colocalized mutations that are transferred together when recombination becomes the dominant force contributing to the success of genetic variants. The y-axis corresponds to x_c/s_b, which quantifies the total amount of clonal interference in the population. The x-axis is a measure of the rarity with which mutations are colocalized and the additional benefit they provide. **(d)** Phase diagram showing different dynamical regimes of adaptation and recombination. The present work focuses on the regime $\log(s_b/r \cdot \tilde{\ell}) \gg 0$ where recombination does not completely break down linkage disequilibrium arising from linked competition.

This solution has a simple interpretation in terms of the processes in Eq. (3). The typical successful beneficial mutation will arise in the high fitness nose of the fitness distribution $x \sim x_c$ (Fig. 2C; middle). Upon surviving the stochastic fluctuations of genetic drift, this mutation will begin to grow deterministically while competing with other mutations in the nose. To take over, the focal mutation will typically survive in the nose for an additional x_c/v_a generations by producing or acquiring a total of $\sim \tilde{k}$ linked mutations (Fig. 2C; top). Each of these mutation and acquisition events will confer a benefit s_b and occur at respective rates U_b and $r \cdot \ell \cdot \tilde{k}$. This implies that the crossover where recombination becomes the dominant force will occur when the transfer and beneficial mutations rates balance $r \cdot \tilde{\ell} \cdot \tilde{k} \sim U_b$, which often corresponds to comparable recombination and mutation rates ($r \sim \mu$).

More rarely, with relative probability $\sim 1/\tilde{k}$, the focal mutation will not take over by staying in the nose. Instead, it will grow exponentially for $\sim x_c/v_a$ generations until it attains an $\mathcal{O}(1)$ frequency near the bulk of the fitness distribution. At a relative rate $r \cdot \tilde{\ell}$, the mutation can then jump back to the nose where it will take over by producing or acquiring linked mutations. This demonstrates how clonal competition allows mutations that would otherwise be purged in an asexual population to attain a high frequency and be rescued by recombination.

For modest population sizes and transfer lengths $(NU_b)^{3/4} \gtrsim L/\tilde{\ell}$, we find that the relative substitution rate will begin to scale with both the total amount of clonal interference NU_b and the fraction of the

genome involved in a recombination event $\tilde{\ell}/L$ (Fig. 1B,C),

$$\log\left(\frac{\theta_r}{\theta_a}\right) \sim (NU_b)^{3/4} \cdot \left(\frac{r\cdot\tilde{\ell}}{U_b}\right)^{\tilde{k}/4-1/8} \cdot \left(\frac{U_b}{s_b}\right)^{1/8} \cdot \left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right) . \tag{6}$$

This solution also has a straightforward interpretation in terms of the processes in Eq. (3). With probability $\sim \tilde{k}\cdot\tilde{\ell}/L$, a new mutation will arise near another preexisting mutation. These mutations will form a colocalized pair since they will typically be transferred together and as a result, their fates will become intertwined (Fig. 3A). The typical successful pair will arise with a moderate fitness disadvantage compared to other mutations in the nose $x_c - s_b \lesssim x_{\text{pair}} \lesssim x_c$ and begin to grow deterministically until it reaches its largest frequency in the bulk. During this time, the pair will have its greatest chance of jumping to a new genetic background, each of which will be randomly drawn from $f(x)$ and obtain a $2s_b$ fitness benefit from the pair. To take over, the pair will need to repeatedly jump back to the nose $\sim \tilde{k}/2$ times until it gets lucky and transfers to an anomalously fit background above x_{pair}. Once there, the pair will transfer to so many backgrounds that it will be guaranteed to sweep through the population. The scaling in Eq. (6) represents a product of the rarity at which these colocalized pairs arise and the compounding $2s_b$ benefit they accumulate with each transfer (Fig. 3B). When this benefit gets sufficiently large, multiple distinct colocalized pairs will attain a large frequency at the same time and recombination will become the dominant force driving the success of genetic variants in the population. We find that the threshold recombination rate where this transition occurs will rapidly decrease with the total genetic diversity, $r\cdot\tilde{\ell} \sim U_b\cdot(U_b/s_b)^{1-1/\tilde{k}}\cdot(L/\tilde{k}\cdot\tilde{\ell})^{2/\tilde{k}}$, which can correspond to extremely small relative recombination rates since $s_b \gg U_b$ (Fig. 1B). In large populations with many competing variants ($\tilde{k} \gg 1$), this threshold rate approaches $r\cdot\tilde{\ell} \sim U_b\cdot U_b/s_b$, which represents an exponentially smaller value than that in small populations. This demonstrates how extremely rare recombination can play a dominant role in evolutionary dynamics.

By extending Eq. (3) to consider contributions from larger numbers of colocalized mutations, we can show that the transition to recombination dominated dynamics can occur at even smaller rates. We find that for extremely large populations $(NU_b) \gtrsim (s_b/U_b)\cdot(L/\tilde{k}\cdot\tilde{\ell})^2$ the dominant mode of fixation will be via the repeated transfer of colocalized triplets (Fig. 3C). In these large populations, predominantly asexual dynamics can breakdown at rates as small as $r\cdot\tilde{\ell} \sim U_b\cdot(U_b/s_b)^{2-2/\tilde{k}}\cdot(L/\tilde{k}\cdot\tilde{\ell})^{2/\tilde{k}}$, highlighting that the role of recombination will continue to increase with the amount of clonal interference in the population (Fig. 3D). In general, the break down of quasi-asexuality will coincide with asymptotically small increases to the rate of adaptation $v_r/v_a - 1 \sim 1/\log(Ns_b)$. While this highlights that recombination will be subtle to detect at the level of the rate of adaptation, our heuristic explanations show that it can play a far more dramatic role under-the-hood influencing the fate of individual mutations. This raises an interesting question. Can we instead turn to the trajectories of successful variants to look for observable signatures of recombination?

Detecting recombination

In large populations, there is a direct connection between the starting fitness of a new mutation and the maximum frequencies it attains in the bulk of the fitness distribution. This allows us to leverage our heuristic explanations above to compare the frequency trajectories of successful transferred variants to mutations that take over in an asexual population. In the absence of recombination, a successful mutation will typically arise in the high-fitness nose of the population $x \sim x_c$ where it will need to produce an additional $\tilde{k}$ linked mutations to take over. Each of these linked mutations will occur at a random time relative to other clones in the nose and begin to grow deterministically until it reaches the

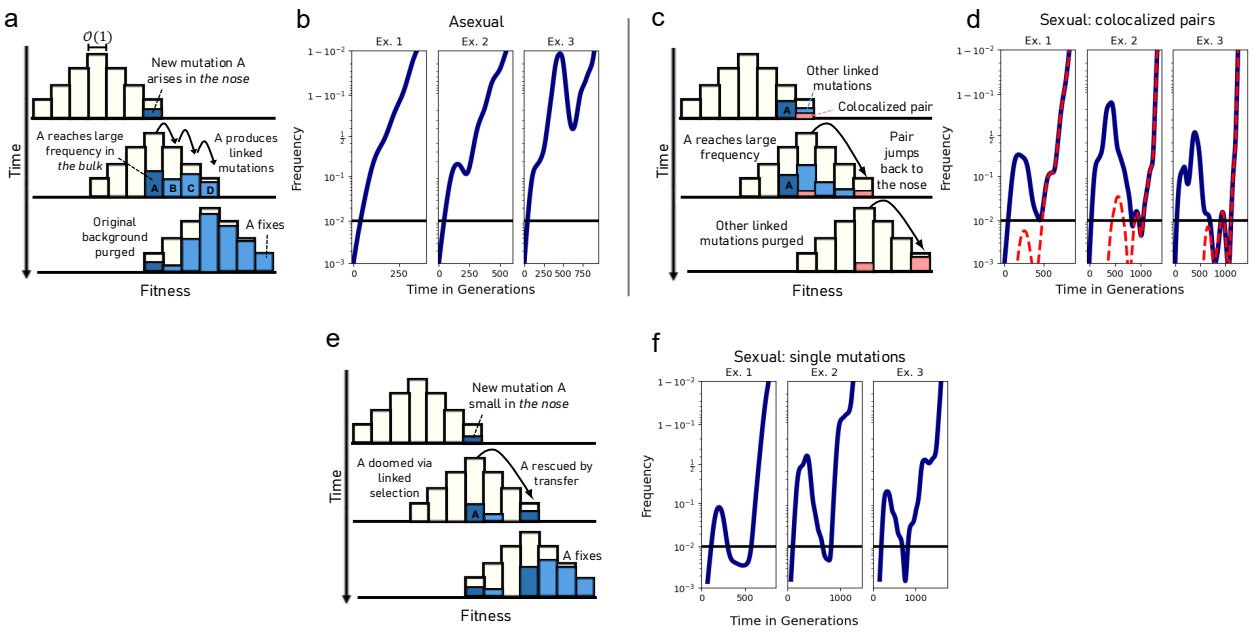


Figure 4: **Trajectories of successful recombined mutations.** **(a)** Schematic of the typical fixation dynamics of a mutation in an asexual population. A successful new mutation arises in the high-fitness nose in the population and reaches $\mathcal{O}(1)$ frequencies in the bulk. It eventually fixes by producing linked mutations that dominate in the nose. **(b)** Typical trajectories of successful mutations that take over in an asexual population. **(c)** Schematic of the typical fixation dynamics of a colocalized pair. The first mutation in the pair arises in the high-fitness nose of the fitness distribution and produces a colocalized mutation. The first mutation reaches large frequencies in the bulk while the colocalized mutation remains small. The colocalized mutations eventually takes over by repeatedly transferring to the nose. **(d)** Typical trajectories of successful colocalized pairs. Solid blue lines denote the first mutation in the pair and red dashed lines denote the second mutation in the pair. **(e)** Schematic of the typical fixation dynamics of a single transferred mutation. The mutation reaches a large overall frequency before its linked descendants are purged. The mutation then rapidly sweeps on its recombined background. **(f)** Typical trajectories of successful single transferred mutations. Simulations have parameters $N = 10^{11}$, $s_b = 5 \cdot 10^{-2}, \tilde{\ell} = 10^4, L = 10^6, \mu = 10^{10}$, $f_b = 10^{-3}$, and $r = 0$ (b) and $r = 10^{-13}$ (d).

bulk (Fig. 4A). As a result, the fixing mutation will typically fluctuate at a large overall frequency $1\% \lesssim f \lesssim 50\%$ before sweeping $T_c \sim x_c/v_a$ generations later (Fig. 4B). Since its chance of taking over will be directly proportional to its observed frequency f (35, 36), a successful mutation will rarely fluctuate below 1%, especially for more than T_c generations.

In stark contrast, we find that this trajectory can be common in successful colocalized pairs. The first mutation in the pair will also typically arise at the nose $x = x_c$ and reach a large overall frequency in the bulk $f \gg 1\%$. At roughly the same time, the colocalized pair will reach the bulk on its initial genetic background; however, due to its initial fitness deficit $x_{\text{pair}} \lesssim x_c$, the pair will remain at small overall frequencies (Fig. 4C). For example, at the rates where recombination starts to appreciably contribute to the rate of fixation, we find that this maximum frequency will be $f \sim (\tilde{\ell}/L)^{1+1/\tilde{k}} \ll 1\%$. As a result, the first mutation in the pair will also drop to these small frequencies within T_c generations where it will stay for an additional time $\sim \tilde{k}/2 \cdot T_c$ until the pair jointly sweeps through the population. This leaves a 'U' like trajectory in genomic sequencing data where a mutation will appear, disappear and then

rapidly sweep (Fig. 4D).

These trajectories are also typical in smaller populations where mutations transfer alone (Fig. 4E). In these scenarios, successful recombined mutations will typically achieve a small but modest overall frequency $f \sim \tilde{k}^{-1}/2 \gtrsim 1\%$ in the bulk and transfer back to the nose. Over the next T_c generations, while this recombined mutation grows to detectable frequencies, the original mutation will decline to a characteristic frequency $f \sim \tilde{k}^{-e}/2 \lesssim 1\%$. After bottoming out, the mutation will rapidly sweep on its new background, leaving a comparatively thinner U-like trajectory (Fig. 4F).

Recombination dominated dynamics

These results, and our present theoretical framework, will remain valid while most mutations take over on their original genetic backgrounds and recombined mutations represent a minority of overall fixations. Once multiple recombined mutations attain a large frequency at the same time, they will compete for dominance preventing any one from quickly taking over. As a result of this competition, we find that successful mutations will arise in short bursts rather than uniformly over time (Fig. 5A). Remarkably, this burstiness drives coherent, periodic oscillations in genetic and fitness diversity, representing a stark departure from the steady-state traveling wave picture in asexual populations (Fig. 5B).

We can understand this behavior most straightforwardly in small populations where recombination dominates $r \cdot \tilde{\ell} \gg U_b$ and successful transfer events typically involve at most one mutation. In these scenarios, new exceptionally fit genotypes at the leading edge of the fitness distribution will continue to drive the population forward. This leads to a key competition in the nose between transferred and new mutations, with the typical ratio of their sizes $\sim r \cdot \tilde{\ell} \cdot k(t)/U_b$ sensitively depending on $k(t)$: the average number of mutations unique to a random acceptor below the nose. When a new burst of successful mutations initially attains a large overall frequency, it will replenish genetic diversity $k \sim \mathcal{O}(1)$ and the next nose class will be dominated by recombinants. As a result, new mutations will compose a minority of the nose and the future donor pool of mutations will be diminished. As the population adapts, the typical age of transferred mutations will grow while existing mutations decline in frequency, depressing both the transfer rate and site frequency spectrum (Fig. 5C). We find that this process will typically continue until the initial burst of new mutations transfers a characteristic number of times $\sim 2\sqrt{x_{cr} \log(r \cdot \tilde{\ell}/U_b)}$ (Fig. 5D). This timescale coincides with a minimum value of the transfer rate, creating a privileged window for a new cohort of successful variants to arise and replenish genetic diversity.

These results suggest that we can think of the nose as advancing with a time-dependent rate. By tracking this effective transfer rate over one typical period of oscillations, we find that the long-term rate of adaption will scale as,

$$v_r \sim 2s_b^2 \frac{\log(Ns_b)}{\log^2(s_b/r \cdot \tilde{\ell})} \cdot \left[1 - \sqrt{\frac{\log(r \cdot \tilde{\ell}/U_b)}{2\log(Ns_b) \cdot \log(s_b/r \cdot \tilde{\ell})}}\right]^2 . \tag{7}$$

This result highlights that the supply of beneficial mutations will continue to limit adaptation, even as recombination becomes the dominant force shaping the fate of genetic variants (Fig. 5F). It also demonstrates that recombination always acts to increase the rate of adaption, reproducing the asexual result in the limit $r \cdot \tilde{\ell} \to U_b$ (Fig. 5E).

Interestingly, in spite of these speed ups, we find that recombination will initially slow down the typical sweep time of genetic variants. This effect emerges from two opposing forces: recombination allows majority frequency mutations to sweep faster, while also extending the lifetime of small frequency mutations (Fig. 5E,F). We can understand both of these these effects as straightforward consequences of our heuristic explanation above. Since successful mutations arise in bursts, the time for a successful mutation to sojourn from one copy to a majority of the population will simply be proportional to the duration of an oscillatory period. As this period grows, recombination can dramatically extend the time for new mutations to attain a majority or be purged from the population. On the other hand, once a mutation attains a majority frequency it will become the typical acceptor background and grow deterministically with each establishment. Since these establishments are comparative quicker in sexual populations $[\propto \log(s_b/r \cdot \tilde{\ell})$ vs. $\log(s_b/U_b)]$, successful mutations will sweep on a faster timescale.

Taken together, these results show that sexual populations exhibit fundamentally non-equilibrium behavior that cannot be captured in the classic traveling wave picture. They also deviate from models of chromosomal reassortment, which give rise to oscillations driven by rare bursts of sex not asexuality (37). As a consequence, there is no simple equivalence principle (e.g. an effective mutation rate or population size) to describe a sexual population using asexual or existing sexual theory. This is often because HGT can have contrasting effects on basic aspects of evolutionary dynamics. Even while alleviating clonal interference and speeding up adaptation, recombination acts to reduce genetic diversity and extend the lifetime of mutations. These results will remain valid while clonal dynamics persist $2\log(Ns_b) \ll \log^2(s_b/r \cdot \tilde{\ell})$ and the typical recombined mutation is not colocalized. In large populations, a similar qualitative picture emerges where the dynamics are characterized by oscillations in the diversity of colocalized pairs.

Discussion

Our results provide a quantitative framework for understanding recombination and adaptation in large clonally evolving populations. We have shown that recombination can become the dominant force driving the success of genetic variants even when it occurs at extremely small rates relative to mutation. These results suggest that small amounts of recombination could play a previously unappreciated role in the evolutionary dynamics of large adapting populations. This could have important consequences for predicting and understanding evolutionary outcomes in natural systems where HGT is widespread like the gut microbiome (38), open ocean (39, 40) and disease-causing pathogens (8, 41–43).

Our theory indicates it may be difficult to detect recombination within a closely related population using traditional metrics like the rate of adaptation or substitution rate. Instead, we show that recombination can leave strong signatures in longitudinal data, potentially opening new avenues to detect its subtle effects. In more distantly related populations where transfer events can be more easily detected, our results suggest that natural selection can strongly bias inferred recombination rates and patterns of linkage. We have shown that this is in part because the role of recombination will depend on the genetic distance between beneficial mutations and the typical lengths of transferred segments. This suggests that future efforts should focus on understanding the relationship between these quantities in specific systems. We have also shown that the role of recombination will strongly depend on factors like the population size and mutation rate. This could influence the regulation of collective behavior like quorum sensing or the evolution of mutation rates.

While we have focused on a simple model where all transferred segments have the same genetic lengths, most organisms can incorporate segments that span a distribution of length scales $\lambda(\ell)$ (Fig. 6A). By extending our theory to consider these scenarios, we find that the distribution of successful segments varies with both the amount of genetic diversity NU_b and genome length L. In small populations where transfers typically involve single mutations, the transfer rate will grow with the ℓ possible start sites and the distribution of successful lengths will have a linear bias $\propto \lambda(\ell) \cdot \ell$. In contrast, for larger populations where successful recombined variants are typically colocalized pairs, we find that the distribution of successful transfer lengths will scale quadratically with ℓ, $\propto \lambda(\ell) \cdot \ell^2$ (Fig. 6B). This scaling arises because successful transfer events can involve any segment longer than the spacing between the colocalized pair $\delta\ell_2$, which suppresses the contribution of smaller segments. These results show that for many shapes of $\lambda(\ell)$ (e.g. an exponential distribution) the fixed distributions will be centered around a characteristic length (Fig. 6B). As a result, our single-ℓ theory can be extended to continuous distributions by defining $\tilde{\ell}$ to coincide with this value. For many simple shapes of $\lambda(\ell)$, $\tilde{\ell}$ will occur near the mean length of a transferred segment $\overline{\ell}$. However, HGT can often involve multiple biological sources and length scales, in which case $\lambda(\ell)$ may have multiple peaks (e.g. a mixture of exponential distributions (Fig. 6C)). In these settings, we find that the average length of a transferred segment can be a poor predictor for what is successful in large populations (Fig. 6D). This demonstrates that selection may strongly bias inferred tract length distributions in empirical data and vary across contexts.

We have also assumed that the fitness benefits of beneficial mutations are additive. In reality, epistatic interactions are often observed in evolving populations (44–47). Our theory allows us to weigh how different features and forms of epistasis interplay with recombination. Interestingly, we find that the effect of pairwise epistasis will strongly depend on the typical distances over which genetic loci interact (Fig. 7A). When this interaction distance exceeds the transfer length, pairwise epistasis can have a negligible effect on the success of recombined mutations (Fig. 7B). This is because mutations will tend to experience a similar average effect across backgrounds. By contrast, when pairwise epistasis acts more locally and interactions fall off beyond a critical distance, the role of recombination can be dramatically enhanced (Fig. 7C). This is because positive interactions can be preserved across backgrounds and exponentially selected as they are repeatedly transferred. Crucially, these enhancements can occur even when the vast majority of interactions are negative (Fig. 7D). This shows that a mutation's average epistatic interaction may be a poor predictor of its long-term success. They also clash with existing theory, which predicts that recombination will inhibit adaptation by breaking apart mutations that interact positively (48). Interestingly, these results also suggest that modular genome architecture, where interacting genes and loci are genetically colocalized, could be under indirect selection in recombining populations. Beyond pairwise interactions, we find that our framework can remain robust to both sign epistasis and diminishing returns epistasis. Even when a large number of beneficial sites reshuffle with each new mutation, recombined variants can still take over at a similar rate (Fig. 7G,H). This is because fixation often occurs over a small number of mutations. Our heuristic descriptions of trajectories allow us to directly estimate these effects. Furthermore, in populations experiencing diminishing returns epistasis, the recombined fraction can remain stable over a large number of substitutions (Fig. 7E,F). This is because recombination's effect scales weakly with the fitness benefits of successful mutations and shifts relatively slowly with the overall number of beneficial sites. As these effects become important over long timescales, our theory reveals how they will impact linkage and genetic diversity along the genome.

We have also assumed that beneficial loci are uniformly distributed along the genome. In reality, beneficial mutations may be concentrated in a small number of hotspot genes n_b (49, 50)(Fig. 7I).

When these target genes are sufficiently rare $n_b \lesssim L/\tilde{\ell}$, the probability that a new mutation is initially colocalized will scale as $\tilde{k}/n_b$. This probability can be dramatically larger than in our simple model, allowing the transition to recombination dominated dynamics to occur at even smaller rates (Fig. 7J).

While our results thus far have largely focused on a simple model where all mutations confer the same fitness benefit, most organisms produce mutations with a continuous range of fitness benefits $\mu(s)$ (Fig. 8A). In these scenarios, previous work has shown that the benefits of mutations at a large frequency in asexual populations will typically occur in a narrow range in the tail of the distribution of fitness effects (34). This allows us to approximate rare recombined mutations as competing against an asexual population with an effective beneficial mutation rate and fitness benefit corresponding to the location and width of this narrow range. This leads to a sensitive interplay between the shape of the DFE, the population size, and the finite number of beneficial sites in the genome (Fig. 8B). For example, when the DFE falls off much faster than exponentially, colocalized pairs will continue to dominate in large populations and both mutations in the pair will have a similar benefit (Fig. 8C). This is because the effective fitness benefit will shift towards the mean where the effective mutation rate resembles the overall beneficial mutation rate. By contrast, when the tail is exponentially distributed, the behavior will depend on the number of beneficial sites and how well the tail is sampled. For example, in small populations $N\mu \ll 1$, the effective benefit will occur near the edge of the fitness distribution x_c and the crossover to recombination dominated dynamics will always involve single transferred mutations (Fig. 8C,D). However, since these driver mutations are extremely rare, this crossover will depend on the population size and occur at asymptotically small rates $r \cdot \tilde{\ell} \sim 1/N$. In larger populations, the magnitude of driver mutations will shift to smaller values well below x_c and colocalized mutations can play an important role in adaptation. This highlights how the shape of the DFE can introduce novel scalings and play a crucial role in adaptive dynamics.

Although we have primarily focused on the competition between beneficial mutations, the overwhelming majority of new variants will typically be deleterious. When these variants have strong costs $|s_d| \gg 1/T_c$, they will typically not reach large enough frequencies to recombine. On the other hand, more weakly deleterious variants $|s_d| \ll 1/T_c$ could hitchhike to larger frequencies. In principle, transfer events could provide an effective benefit to a genotype by overwriting these variants. However, given empirical estimates of the deleterious mutation rates, these variants will not be sufficiently numerous at large frequency to affect adaptive dynamics.

In this work, we have also focused on horizontal gene transfer within closely related populations. However, natural populations typically import homologous segments from more diverged strains, where multiple barriers limit the evolutionary impact of introgressed DNA (38). For example, diverged segments may be imported at lower overall rates, have smaller typical lengths, or be more deleterious on average. Our results suggest that these differences can exponentially suppress the fixation rate of introgressed fragments relative to de novo mutations and offer a framework to assess the relative importance of these barriers. Moreover, introgressed fragments may contain a mixture of beneficial and deleterious mutations that can be unlinked over repeated transfers. Our work shows that this process will depend on a fragment's initial fitness benefit and the overall frequencies it attains, as well as the future benefits of unlinking deleterious mutations.

Finally, we have focused on a simple model of localized recombination occurring at roughly constant rates. However, certain viral and phage and viral species can engage in alternative mechanisms of recombination, including chromosomal reassortment and template-switching. Previous work has shown that rare reassortment can dramatically speed up adaptation (37). Another body of work has

studied recombination and adaptation, but has focused on frequent crossover-like recombination (51–53). These models lead to a “linkage block” picture of evolution wherein local segments of genome evolve effectively asexually, but are otherwise unlinked to distal regions. When crossover-like recombination is rare, our theoretical approach could be extended to predict when predominantly asexual dynamics break down. More generally, recombination rates can also evolve over time. This can involve trade-offs in direct opportunity costs and indirect benefits of bringing together beneficial combinations of mutations. Our work provides a promising theoretical framework to understand how natural selection acts on these trade-offs in adapting populations.

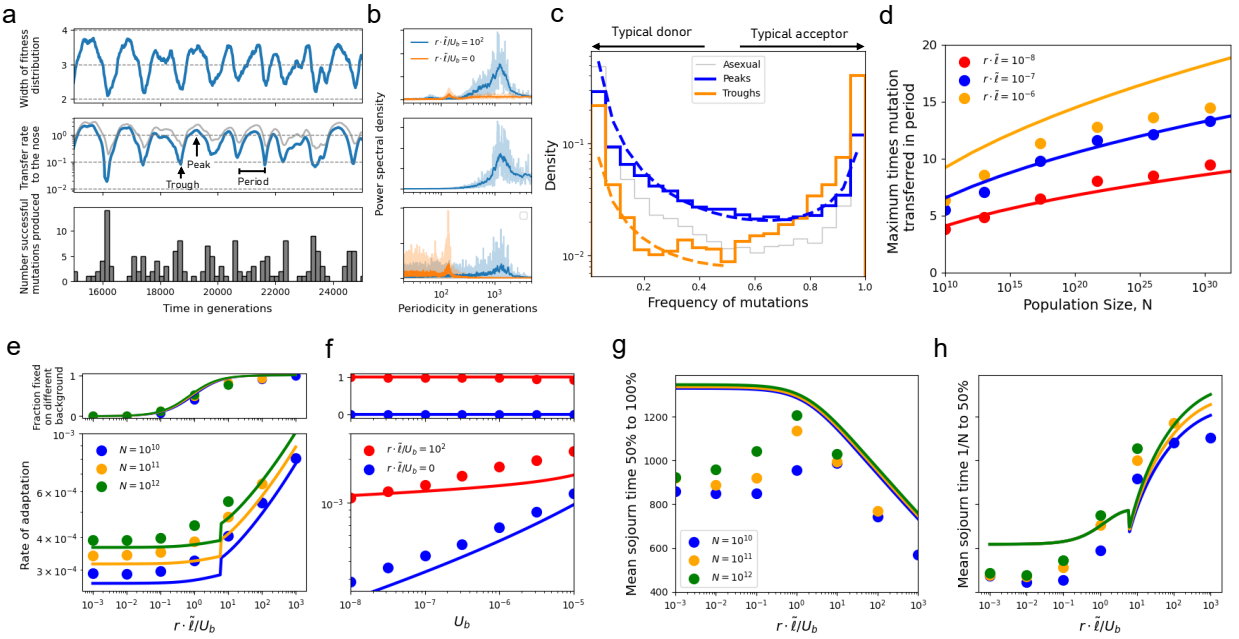


Figure 5: **Recombination dominated dynamics .** **(a)** Coherent periodic oscillations in the width of the fitness distribution, transfer rate of mutations to the nose, and number of successful mutations produced over time. The gray line denotes the average pairwise heterozygosity. **(b)** Power spectral density analysis of $\sim 10^5$ generations of data from panel a) shows concentration around characteristic frequency across multiple metrics of fitness and genetic diversity. **(c)** Average site frequency spectrum of mutations at time points corresponding to peak and trough transfer rates from panel a). Mutations at a minority frequency represent the typical donated mutation, while mutations at a majority represent the typical donor background. **(d)** The maximum number of transfers in a period from quasi-deterministic simulations of large population sizes. In these simulations, only genotypes in the nose are allowed to mutate and recombine, while donors are sampled randomly from the population. **(e,f)** Fraction of mutation fixed on different background (top) and rate of adaptation (bottom). **(g)** Mean sojourn time from a majority frequency to fix. **(h)** Mean sojourn time from one copy to a majority frequency. Deterministic simulations in d) have parameters $s_b = 10^{-2}, \tilde{\ell} = 10^3, L = 10^6, \mu = 10^{-13}$, and $f_b = 10^{-3}$. Full simulations have parameters $s_b = 5 \cdot 10^{-2}, \tilde{\ell} = 10^3, L = 10^6, \mu = 10^{-10}$, and $f_b = 10^{-3}$. In all simulations the distance between beneficial sites is fixed at $\tilde{\ell} - 1$ to allow only the transfer of individual mutations.

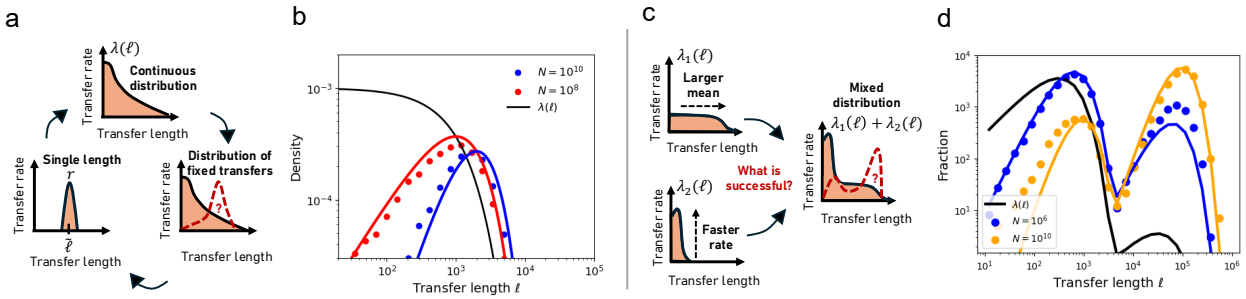


Figure 6: **Distributions of transfer lengths . (a)** The simple model with a single transfer length $\tilde{\ell}$ can be extended to consider a continuous distribution of transfer lengths. The red dashed line denote the distribution of transfer lengths that contribute to the fixation of successful recombined mutations. **(b)** Distribution of fixed transfer lengths given an exponential distribution of transfer lengths. The black solid line denotes $\lambda(\ell) \equiv 1/\overline{\ell} \cdot e^{-\ell/\overline{\ell}}$. **(c)** An alternative model of a distribution of transfer lengths that is a mixture of a distribution with a smaller typical length and faster rate and a distribution with a larger typical length and slower rate. **(d)** A concrete example of the scenario in panel c): a mixed distribution of two exponential distributions $\lambda(\ell) \equiv f_1 \cdot 1/\overline{\ell_1} \cdot e^{-\ell/\overline{\ell_1}} + (1-f_1) \cdot 1/\overline{\ell_2} \cdot e^{-\ell/\overline{\ell_2}}$. Transfers from λ_1 occur 10^3 more frequently and correspond to a transfer rate 10 times larger. Symbols denote the results of forward time simulations. Solid colored lines denote the theoretical predictions. The distribution of transfer lengths for a distribution of transfer lengths with two lengths $\tilde{\ell}_1 = 3.16 \cdot 10^2$ and $\tilde{\ell}_2 = 10^4$ and two rates $r_1 = 10^{-11}$ and $r_2 = 3.16 \cdot 10^{-13}$. Simulations have parameters $U_b = 10^{-6}, s_b = 5 \cdot 10^{-2}, r = 10^{-11}, L = 10^6, f_b = 10^{-4}$, and $f_1 = 1 - 10^{-3}$

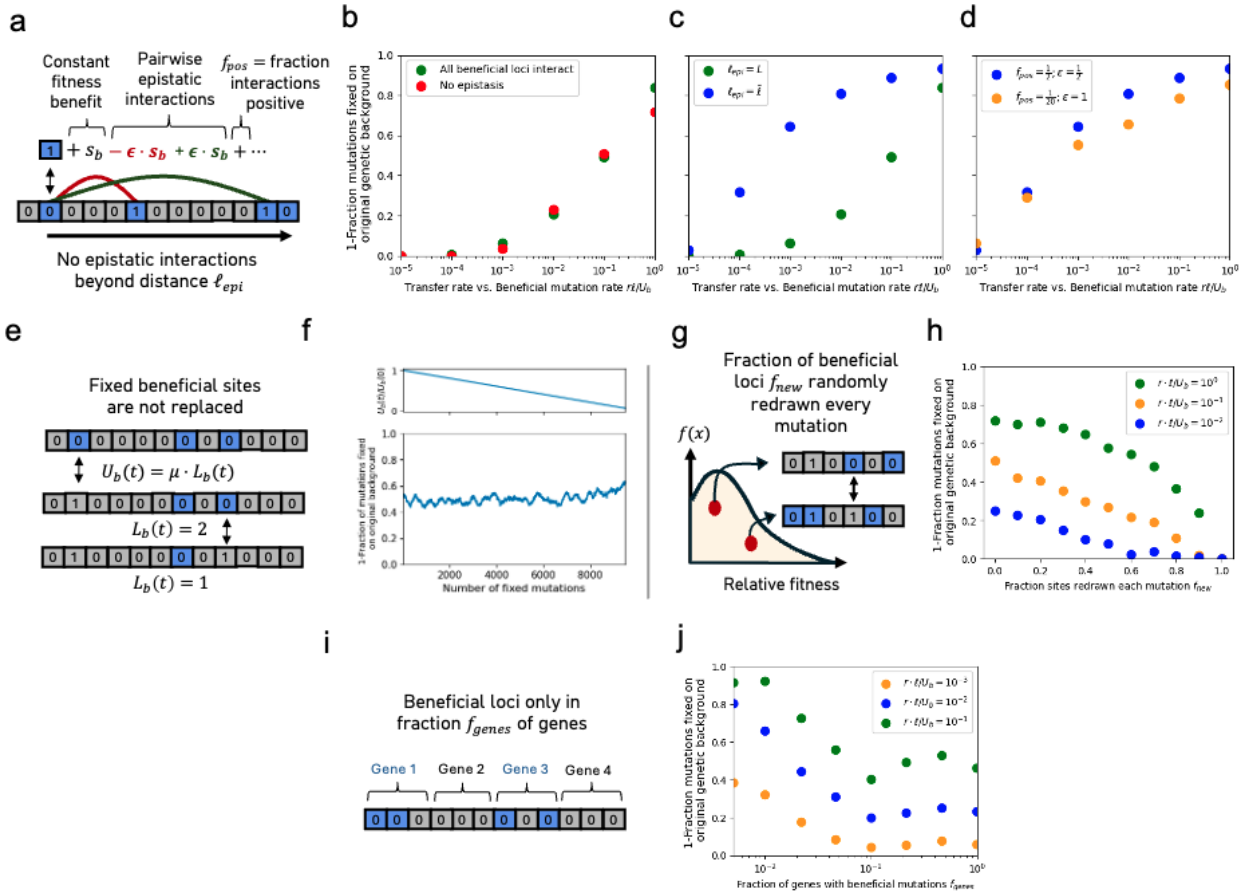


Figure 7: **Epistasis and recombination. (a)** A simple model of pairwise epistasis where all $L \cdot f_b$ beneficial sites confer the same first order benefit s_b and each of the $(L \cdot f_b)^2$ pairwise interactions between beneficial sites randomly confers an additional benefit (or cost) of ϵ, where f_{pos} is the probability that any given interaction is positive. The total fitness benefit of a site i arises as a sum between these two terms $s_i = s_b + \epsilon \cdot \sum_{j \neq i} s_b \cdot g_i \cdot g_j$, where all sites are initially $g_i = 0$ and become $g_i = 1$ if mutated. **b)** The fraction of mutations that take over on their original background for the model in Fig.1B with no epistatic effects and the pairwise model where $\epsilon = \frac{1}{2}$ and $f_{pos} = \frac{1}{2}$. When pairwise epistasis has an equal chance of being positive or negative, a mutations fitness benefit is dominated by first order effects. **c)** The fraction of mutations that take over on their original background increases for a model of local vs. global pairwise epistasis. Pairwise interactions for loci separated by a distance greater than ℓ_{epi} are set to zero. When this distance is smaller than the transfer length $\tilde{\ell}$, pairwise benefits can be preserved across multiple backgrounds. As a result recombination can play a role at smaller overall rates. **d)** The fraction of mutations that take over on their original background for the model in Fig.1B with varying tradeoffs in the strength of epistatic interactions and fraction of interactions that are positive. Rare, larger effect interactions can have a similar effect on the substitution rate compared to weaker interactions that are more likely to be positive. **e)** A simple model of diminishing returns epistasis where beneficial loci are not randomly replaced after fixing. **f)** The fraction of mutations that take over on their original background. A large number of beneficial sites can fix while exerting only a small effect on the role of recombination. **g)** A simple maximally epistatic model where a fraction of the beneficial sites available to a genotype are randomly redrawn from all loci every mutation. **h)** The fraction of mutations fixed on their original background. Even if a large fraction (e.g. 50%) of beneficial sites are redrawn each mutation, this fraction can remain relatively constant. **i)** A simple model of non-uniform distributions of beneficial loci along the genome. Beneficial sites are concentrated in a small fraction of genes. **j)** As beneficial mutations become more clustered, colocalized mutations will be produced at a larger rate, allowing recombination to become important at smaller overall rates.

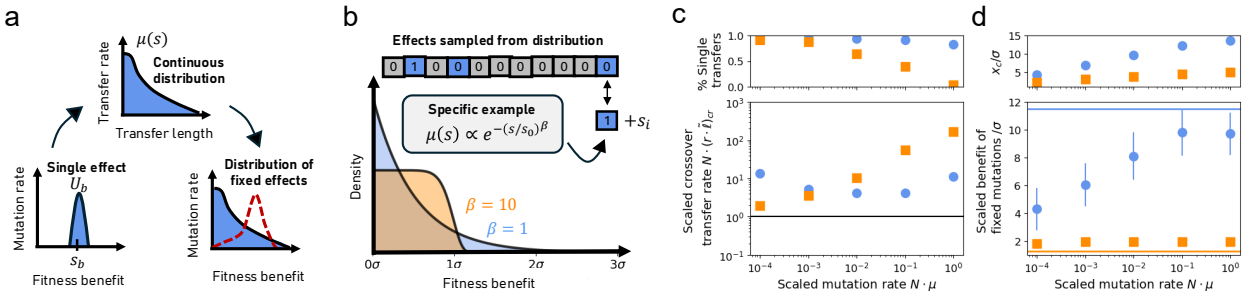


Figure 8: **Continuous distributions of fitness effects. (a)** The simple model where all mutations have the same benefit s_b can be extended to consider a continuous distribution of fitness effects. The red dashed line denote the distribution of fixed beneficial mutations. In the quasi-asexual limit, successful mutations tend to concentrate in a narrow region in the tail of the distribution. **(b)** An extension of the simple model where the fitness benefits of beneficial loci are sampled from a stretched exponential distribution with shape and scale parameters β and s_0. When $\beta = 1$, the distribution simplifies to an exponential shape. In the opposite limit where $\beta \gg 1$, the distribution falls off sharply above s_0. The displayed distributions both have the same average benefit σ. **(c)** The scaled crossover recombination rate where recombined mutations fix on a par with those on their original background $\theta_a = \theta_r$ (bottom). The fraction of successful transfers involving single mutations (top). When $\beta = 1$ and $N\mu \lesssim 1$, the scaled recombination rate approaches the inverse population size $(r \cdot \tilde{\ell})_{cr} \sim 1/N$ and most transfers involve a single mutation. This demonstrates recombination can become important at asymptotically small rates. For larger values of β, colocalized pairs become dominant at moderate population sizes $NU_b \approx L/\tilde{\ell}$ and the crossover rate scales with $\log(s_0/U_b)$. **(d)** The scaled fitness benefit and standard deviation of successful mutations at the crossover recombination rate (bottom). The average width of the fitness distribution normalized by the average benefit produced. When $\beta = 1$, the typical successful benefit will approach the maximum sampled benefit (solid line). For larger values of β, the typical successful benefit remains near the mean and the width of the fitness benefit is proportionally smaller. Simulations have parameters $N = 10^7 - 10^{11}$, $U_b = 10^{-7}$, $s_b = 2 \cdot 10^{-2}$, $r = 10^{-10} - 10^{-15}$, $\tilde{\ell} = 10^4$, $L = 10^6$, $f_b = 10^{-5}$, $\beta = 1$ and $\beta = 10$.

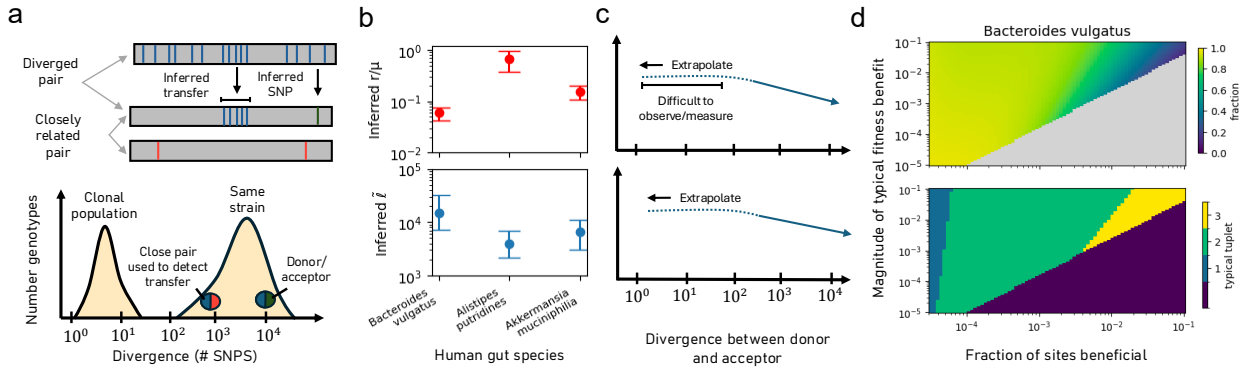


Figure 9: **Homologous recombination in empirical example. (a)** A typical method of inferring recombination rates and length scales from empirical data. Tracts of atypically high heterozygosity between closely related individuals evidence recent transferred between one individual in the pair and a more diverged genotype. **(c)** Inferred recombination rates and length scales from three example human gut species from (38). **(c)** While it is difficult to observe recombination in more closely related populations, it may be reasonable to extrapolate inferred rates and length scales. **(d)** Theoretical predictions for the fraction of mutations taking over on a new genetic background (top) and the typical tuplet size of successful recombined mutations (bottom) for empirical estimates of Bacteroides vulgatus: N=10^{11}, $\mu = 10^{-10}$, $L = 5 \cdot 10^6$, $\tilde{\ell} \approx 1.5 \cdot 10^4$, and $r/\mu \approx 8 \cdot 10^{-2}$.

References

[1] Shannon M Soucy, Jinling Huang, and Johann Peter Gogarten. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 16:472–482, 2015.

[2] Mathieu Groussin, Mathilde Poyet, Ainara Sistiaga, Sean M Kearney, Katya Moniz, Mary Noel, Jeff Hooker, Sean M Gibbons, Laure Segurel, Alain Froment, Rihlat Said Mohamed, Alain Fezeu, Vanessa A Juimo, Sophie Lafosse, Francis E Tabe, Catherine Girard, Deborah Iqaluk, Le Thanh Tu Nguyen, B Jesse Shapiro, Jenni Lehtimäki, Lasse Ruokolainen, Pinja P Kettunen, Tommi Vatanen, Shani Sigwazi, Audax Mabulla, Manuel Domínguez-Rodrigo, Yvonne A Nartey, Adwoa Agyei-Nkansah, Amoako Duah, Yaw A Awuku, Kenneth A Valles, Shadrack O Asibey, Mary Y Afihene, Lewis R Roberts, Amelie Plymoth, Charles A Onyekwere, Roger E Summons, Ramnik J Xavier, and Eric J Alm. Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell*, 184:2053–2067.e18, 4 2021. doi: 10.1016/j.cell.2021.02.052.

[3] Alyssa G Kent, Albert C Vill, Qiaojuan Shi, Michael J Satlin, and Ilana Lauren Brito. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial hi-c. *Nature Communications*, 11:4379, 2020.

[4] Christopher M Thomas and Kaare M Nielsen. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3:711–721, 2005.

[5] William P Hanage. Not so simple after all: Bacteria, their population genetics, and recombination. *Cold Spring Harbor Perspectives in Biology*, 8, 7 2016. 10.1101/cshperspect.a018069.

[6] Tim F Cooper. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of escherichia coli. *PLOS Biology*, 5(9):1–7, 08 2007.

[7] R A Neher, B I Shraiman, and D S Fisher. Rate of adaptation in large sexual populations. *Genetics*, 184:467–481, 2 2010.

[8] William Paul Hanage, Christophe Fraser, Jing Tang, Thomas Richard Connor, and Jukka Corander. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science*, 324:1454–1457, 6 2009. doi: 10.1126/science.1171908.

[9] An N T Nguyen, Rebecca Gorrell, Terry Kwok, Tim Connallon, and Michael J McDonald. Horizontal gene transfer facilitates the molecular reverse-evolution of antibiotic sensitivity in experimental populations of h. pylori. *Nature Ecology Evolution*, 8:315–324, 2024.

[10] L D Bowler, Q Y Zhang, J Y Riou, and B G Spratt. Interspecies recombination between the pena genes of neisseria meningitidis and commensal neisseria species during the emergence of penicillin resistance in n. meningitidis: natural events and laboratory simulation. *Journal of Bacteriology*, 176:333–337, 1 1994. doi: 10.1128/jb.176.2.333-337.1994.

[11] G Laible, B G Spratt, and R Hakenbeck. Interspecies recombinational events during the evolution of altered pbp 2x genes in penicillin-resistant clinical isolates of streptococcus pneumoniae. *Molecular Microbiology*, 5:1993–2002, 8 1991.

[12] Nelson Frazão, Ana Sousa, Michael Lässig, and Isabel Gordo. Horizontal gene transfer overrides mutation in escherichia coli colonizing the mammalian gut. *Proceedings of the National Academy of Sciences*, 116(36):17906–17915, 2019.

[13] Shai Slomka, Itamar Françoise, Gil Hornung, Omer Asraf, Tammy Biniashvili, Yitzhak Pilpel, and Orna Dahan. Experimental evolution of bacillus subtilis reveals the evolutionary dynamics of horizontal gene transfer and suggests adaptive and neutral effects. *Genetics*, 216(2):543–558, 10 2020.

[14] Nicholas J. Croucher, Simon R. Harris, Christophe Fraser, Michael A. Quail, John Burton, Mark van der Linden, Lesley McGee, Anne von Gottberg, Jae Hoon Song, Kwan Soo Ko, Bruno Pichon, Stephen Baker, Christopher M. Parry, Lotte M. Lambertsen, Dea Shahinas, Dylan R. Pillai, Timothy J. Mitchell, Gordon Dougan, Alexander Tomasz, Keith P. Klugman, Julian Parkhill, William P. Hanage, and Stephen D. Bentley. Rapid pneumococcal evolution in response to clinical interventions. *Science*, 331(6016):430–434, 2011.

[15] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetical Research*, 23:23–35, 1974.

[16] B Charlesworth, P H Harvey, and N H Barton. Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355:1553–1562, 11 2000. doi: 10.1098/rstb.2000.0716.

[17] H J Muller. Some genetic aspects of sex. *The American Naturalist*, 66:118–138, 1932.

[18] Natalja Strelkowa and Michael Lässig. Clonal interference in the evolution of influenza. *Genetics*, 192:671–682, 10 2012.

[19] Fabio Zanini, Johanna Brodin, Lina Thebo, Christa Lanz, Göran Bratt, Jan Albert, and Richard A Neher. Population genomics of intrapatient hiv-1 evolution. *eLife*, 4:e11282, 2015.

[20] Nash D Rochman, Yuri I Wolf, Guilhem Faure, Pascal Mutz, Feng Zhang, and Eugene V Koonin. Ongoing global and regional adaptive evolution of sars-cov-2. *Proceedings of the National Academy of Sciences*, 118:e2104241118, 7 2021. doi: 10.1073/pnas.2104241118.

[21] João Barroso-Batista, Ana Sousa, Marta Lourenço, Marie-Louise Bergman, Daniel Sobral, Jocelyne Demengeot, Karina B Xavier, and Isabel Gordo. The first steps of adaptation of escherichia coli to the gut are dominated by soft sweeps. *PLOS Genetics*, 10:e1004182–, 3 2014.

[22] Daniel P G H Wong and Benjamin H Good. Quantifying the adaptive landscape of commensal gut bacteria using high-resolution lineage tracking. *bioRxiv*, page 2022.05.13.491573, 1 2023.

[23] Richard A Neher. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 44:195–215, 11 2013. doi: 10.1146/annurev-ecolsys-110512-135920.

[24] Ron Sender, Shai Fuchs, and Ron Milo. Revised estimates for the number of human and bacteria cells in the body. *PLOS Biology*, 14:e1002533–, 8 2016.

[25] Olivia M Ghosh and Benjamin H Good. Emergent evolutionary forces in spatial models of luminal growth and their application to the human gut microbiota. *Proceedings of the National Academy of Sciences*, 119:e2114931119, 7 2022. doi: 10.1073/pnas.2114931119.

[26] Isabel Gordo. Evolutionary change in the human gut microbiome: From a static to a dynamic view. *PLOS Biology*, 17(2):1–5, 02 2019.

[27] W G Hill and Alan Robertson. The effect of linkage on limits to artificial selection. *Genetical Research*, 8:269–294, 1966.

[28] Elisheva Cohen, David A Kessler, and Herbert Levine. Recombination dramatically speeds up evolution of finite populations. *Physical Review Letters*, 94:98102, 3 2005.

[29] Nandita R Garud, Benjamin H Good, Oskar Hallatschek, and Katherine S Pollard. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLOS Biology*, 17:e3000102–, 1 2019.

[30] Shijie Zhao, Tami D Lieberman, Mathilde Poyet, Kathryn M Kauffman, Sean M Gibbons, Mathieu Groussin, Ramnik J Xavier, and Eric J Alm. Adaptive evolution within gut microbiomes of healthy people. *Cell Host I& Microbe*, 25:656–667.e8, 5 2019. doi: 10.1016/j.chom.2019.03.007.

[31] Zhiru Liu and Benjamin H Good. Dynamics of bacterial recombination in the human gut microbiome. *PLOS Biology*, 22:e3002472–, 2 2024.

[32] James T. Ferrare and Benjamin H. Good. Evolution of evolvability in rapidly adapting populations. *Nature Ecology & Evolution*, 8(11):2085–2096, Nov 2024.

[33] Michael M Desai and Daniel S Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176:1759–1798, 7 2007.

[34] Benjamin H Good, Igor M Rouzine, Daniel J Balick, Oskar Hallatschek, and Michael M Desai. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences*, 109:4950–4955, 3 2012. doi: 10.1073/pnas.1119910109.

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

[36] R A Neher and B I Shraiman. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188:975–996, 8 2011.

[37] Michael T Pearce and Daniel S Fisher. Rapid adaptation in large populations with very rare sex: Scalings and spontaneous oscillations. *Theoretical Population Biology*, 129:18–40, 2019.

[38] Zhiru Liu and Benjamin H Good. Dynamics of bacterial recombination in the human gut microbiome. *bioRxiv*, page 2022.08.24.505183, 1 2022.

[39] Lauren D. McDaniel, Elizabeth Young, Jennifer Delaney, Fabian Ruhnau, Kim B. Ritchie, and John H. Paul. High frequency of horizontal gene transfer in the oceans. *Science*, 330(6000):50–50, 2010.

[40] Ying Sun and Haiwei Luo. Homologous recombination in core genomes facilitates marine bacterial adaptation. *Applied and Environmental Microbiology*, 84(11):e02545–17, 2018.

[41] Sandip Paul, Elena V. Linardopoulou, Mariya Billig, Veronika Tchesnokova, Lance B. Price, James R. Johnson, Sujay Chattopadhyay, and Evgeni V. Sokurenko. Role of homologous recombination in adaptive diversification of extraintestinal escherichia coli. *Journal of Bacteriology*, 195(2):231–242, 2013.

[42] Hannah D. Steinberg and Evan S. Snitkin. Homologous recombination in clostridioides difficile mediates diversification of cell surface features and transport systems. *mSphere*, 5(6):10.1128/msphere.00799–20, 2020.

[43] Chrispin Chaguza, Madikay Senghore, Ebrima Bojang, Rebecca A. Gladstone, Stephanie W. Lo, Peggy-Estelle Tientcheu, Rowan E. Bancroft, Archibald Worwui, Ebenezer Foster-Nyarko, Fatima Ceesay, Catherine Okoi, Lesley McGee, Keith P. Klugman, Robert F. Breiman, Michael R. Barer, Richard A. Adegbola, Martin Antonio, Stephen D. Bentley, and Brenda A. Kwambana-Adams. Within-host microevolution of streptococcus pneumoniae is rapid and adaptive during natural colonisation. *Nature Communications*, 11(1):3442, Jul 2020.

[44] Benjamin H Good and Michael M Desai. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics*, 199:177–190, 1 2015.

[45] Daniel J. Kvitek and Gavin Sherlock. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLOS Genetics*, 7(4):1–11, 04 2011.

[46] Hsin-Hung Chou, Nigel F. Delaney, Jeremy A. Draghi, and Christopher J. Marx. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLOS Genetics*, 10(2):1–11, 02 2014.

[47] Aisha I. Khan, Duy M. Dinh, Dominique Schneider, Richard E. Lenski, and Tim F. Cooper. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034):1193–1196, 2011.

[48] Sarah Perin Otto and Nick H Barton. The evolution of recombination: Removing the limits to natural selection. *Genetics*, 147(2):879–906, 10 1997.

[49] Olivier Tenaillon, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. The molecular diversity of adaptive convergence. *Science*, 335(6067):457–461, 2012.

[50] Benjamin H Good, Michael J McDonald, Jeffrey E Barrick, Richard E Lenski, and Michael M Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551:45–50, 2017.

[51] Daniel B Weissman and Oskar Hallatschek. The rate of adaptation in large sexual populations with linear chromosomes. *Genetics*, 196:1167–1183, 4 2014.

[52] Richard A Neher, Taylor A Kessinger, and Boris I Shraiman. Coalescence and genetic diversity in sexual populations under selection. *Proceedings of the National Academy of Sciences*, 110:15836–15841, 9 2013. doi: 10.1073/pnas.1309697110.

[53] Daniel B Weissman and Nicholas H Barton. Limits to the rate of adaptive substitution in sexual populations. *PLOS Genetics*, 8:e1002740–, 6 2012.

[54] Dmitry A Kondrashov and Fyodor A Kondrashov. Topological features of rugged fitness landscapes in sequence space. *Trends in Genetics*, 31:24–33, 2015.

[55] Katya Kosheleva and Michael M Desai. The dynamics of genetic draft in rapidly adapting populations. *Genetics*, 195(3):1007–1025, 11 2013.

[56] Daniel S Fisher. Asexual evolution waves: fluctuations and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2013:P01011, 2013.

[57] Matthew J Melissa, Benjamin H Good, Daniel S Fisher, and Michael M Desai. Population genetics of polymorphism and divergence in rapidly evolving populations. *Genetics*, 221:iyac053, 8 2022.

[58] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

Supplementary Material

Contents

Page

1 General evolutionary dynamics with horizontal gene transfer 4

2 Independent sites limit 5

3 Rare beneficial sweeps 5

3.1 Sweeps against a clonal background 6

3.2 Initial breakdown of clonal sweeps 6

3.3 Deleterious mutations 7

4 Clonal interference and recombination 7

5 Fast recombination limit 10

6 Asexual limit 12

7 Quasi-asexual theory 14

7.1 Contributions to the branching process recursion for the fixation probability 15

7.2 Branching process recursion with recombination 17

7.3 Solution for the branching process recursion 18

8 Solution for a simple model of distribution of transfer lengths 19

8.1 Mutations that are not initially colocalized 20

8.1.1 Single mutation acquisition limit $\eta \lesssim 1$ 22

8.1.2 Beyond the single mutation acquisition limit 23

8.2 Mutations that are initially colocalized 24

8.3 Assessing the dominant contribution from recombination 28

8.4 Mutations arising as colocalized triplets . . . 30
8.4.1 Single transfer limit . . . 30
8.4.2 Repeated transfers of colocalized triplets . . . 32
8.4.3 Breakdown of asexuality . . . 32
8.5 Larger tuplet sizes . . . 33
8.6 Rate of adaptation . . . 33

9 Asexual genetic diversity and evolutionary trajectories 34

9.1 Stochastic dynamics in the nose . . . 35
9.2 Fate of individual variants: linked selection . . . 36
9.3 Linked selection: jackpotting and busting . . . 37
9.4 Linked selection: genetic draft and genetic drowning . . . 39
9.5 Trajectories of successful mutations . . . 39
9.6 Sojourn times . . . 40

10 Recombination dominated dynamics: single transfer regime 41

10.1 Stochastic dynamics In the nose . . . 41
10.2 Forward-time dynamical approach . . . 42
10.3 Initial transfer rate . . . 43
10.4 Fate of individual mutations . . . 43
10.5 First update of the site frequency spectrum . . . 44
10.6 First update to the transfer rate . . . 47
10.7 General updates of the transfer rate and site frequency spectrum . . . 48
10.8 Depletion and exhaustion of genetic diversity . . . 48
10.9 Replenishment of genetic diversity and steady state dynamics . . . 51
10.10 Upper bound on the period . . . 51
10.11 Long-term estimates of the rate of adaptation . . . 52

10.12 Duration of period . . . 52
10.13 Sojourn times . . . 53
10.14 Regime of validity . . . 54

11 Recombination dominated dynamics: colocalized transfer regime 54

11.1 Branching process dynamics in the nose . . . 54
11.2 Initial sexual dynamics . . . 56
11.3 Time for colocalized pairs to reach large frequency . . . 57
11.4 Oscillations in the nose over short timescales . . . 57
11.5 Unstable non-equilibrium oscillations in the nose . . . 58
11.6 Width of the fitness distribution . . . 59
11.7 Exhausting colocalized pairs over long timescales . . . 59

12 Distributions of transfer lengths 61

13 Distributions of Fitness Effects 62

13.1 Specific distributions of fitness effects . . . 64

14 Incorporating deleterious mutations 65

15 More general models of epistasis and genome organization 66

15.1 Diminishing returns epistasis . . . 66
15.2 Non-uniform distributions of beneficial loci . . . 67
15.3 Pairwise epistasis . . . 68
15.4 Uncorrelated beneficial loci . . . 68

16 Numerical Methods 69

16.1 Forward time simulations . . . 69
16.2 Theoretical predictions . . . 69

1 General evolutionary dynamics with horizontal gene transfer

We consider a well-mixed population of N individuals. The dynamics of a new mutation can be written in the general form,

$$\frac{\partial f(\vec{g})}{\partial t} = \underbrace{\left[X(\vec{g}) - \overline{X}(t)\right] f(\vec{g})}_{\text{selection}} + \underbrace{\sum_{\vec{g}'} M(\vec{g}' \to \vec{g}) f(\vec{g}') - \sum_{\vec{g}} M(\vec{g} \to \vec{g}') f(\vec{g})}_{\text{mutation}}$$
$$+ \underbrace{\sum_{\vec{g}'} R(\vec{g}' \to \vec{g}) f(\vec{g}') - \sum_{\vec{g}} R(\vec{g} \to \vec{g}') f(\vec{g})}_{\text{recombination}} + \underbrace{\sqrt{\frac{f(\vec{g})}{N}}\eta(\vec{g}, t) - f(\vec{g}) \sum\nolimits_{\vec{g}'} \sqrt{\frac{f(\vec{g}')}{N}}\eta(\vec{g}', t)}_{\text{genetic drift}}, \tag{8}$$

where $f(\vec{g}, t)$ is the frequency of genotype $\vec{g}$, $X(\vec{g})$ is the (log) fitness of genotype $\vec{g}$, $\overline{X}(t) = \sum_{\vec{g}} X(\vec{g}) f(\vec{g}, t)$ is the mean fitness of the population, $R(\vec{g} \to \vec{g}')$ and $M(\vec{g} \to \vec{g}')$ are the recombination and mutation rates from genotype $\vec{g}$ to $\vec{g}'$, and $\eta(\vec{g}, t)$ is an uncorrelated Brownian noise term. We initially consider a simple bi-allelic genome with $L \gg 1$ genetic loci indexed by their location on the genome,

$$\vec{g} \equiv \{g_1, g_2, \ldots g_L\}\,, \tag{9}$$

where $g_i = 1$ if there is a mutation at site i and 0 otherwise. The new mutation will arise at one of these sites and compete to take over.

A central contribution of this work is to understand how recombination influences this competition. Recombination reshuffles genetic variation across competing backgrounds, generating new genotypes from existing variation. This process can can occur through various physical mechanisms with implications for how and which genotypes can be generated. In this work, we focus on recombination mediated by horizontal gene transfer in microbial populations.

A recombination event occurs when a random donor genome transfers a genetic segment of length ℓ to a random acceptor genome, which then integrates the segment, overwriting any existing variants. The rates and lengths of transferred segments are drawn from the distribution of transfer lengths $R(\ell|\vec{g})$. In principle, this distribution can vary with the genotype. This effect has been extensively studied ('modifier theory' (48)) and has important implications for the evolution of recombination. To tease apart the basic role of recombination in adaptation, we relax this genotype dependence and consider a constant distribution of transfer lengths $R(\ell) \equiv r \cdot \lambda(\ell)$, where r is the per base pair recombination rate and $\lambda(\ell)$ is a normalized probability density function. The total transfer rate of a focal site is simply $r \cdot \overline{\ell}$, where $\overline{\ell} \equiv \int \ell \cdot \lambda(\ell) d\ell$ is the mean transfer length. Consistent with the localized nature of HGT, we assume the mean transfer length remains much smaller than the total genome length $\overline{\ell} \ll L$. This assumption distinguishes recombination mediated by horizontal gene transfer from classic meiotic recombination. Although both processes can unlink mutations from their genetic background, transfer events will typically preserve linkage on an acceptor genome. As we will show, this will allow us to dramatically simplify the recombination kernel in Eq. (8) in our scenarios of interest.

Naturally, the role of recombination will depend on the degree of genetic diversity within the population. In turn, this diversity will be strongly influenced by the fitness landscape $X(\vec{g})$. There are many possible fitness landscapes one could consider. These are often parameterized as a power series

involving different combinations of loci,

$$X(\vec{g}) = \underbrace{\sum_{i=1}^{L} s_i g_i}_{\text{'first-order' epistasis}} + \underbrace{\sum_{i<i'} \epsilon_{i,i'} g_i g_{i'}}_{\text{'second-order' epistasis}} + \underbrace{\sum_{i<i'<i''} \epsilon_{i,i',i''} g_i g_{i'} g_{i''}}_{\text{'third-order' epistasis}} + \dots \tag{10}$$

where nonzero values of the $\epsilon_{i,\dots,i'}$ coefficients correspond to epistatic interactions between loci (54). Following previous work (44), we refer to these interactions as *microscopic epistasis* since they capture the detailed dependencies of individual loci. While these dependencies can be arbitrarily complicated, the evolutionary dynamics are known to simplify in several asymptotic limits involving the recombination kernel and fitness landscape.

2 Independent sites limit

One particularly well-studied limit arises when recombination occurs extremely fast. In these scenarios, mutations are rapidly unlinked from their original genetic background. As a result, each site in the genome can be approximated as evolving independently. This allows us to neglect the potentially complicated details of microscopic epistasis and describe the fate of a mutation in terms of its first order effect s_i,

$$\frac{df}{dt} = \underbrace{f \cdot (1-f) \cdot s_i}_{\text{selection}} + \underbrace{(1-f) \cdot \mu_i}_{\text{forward mutation}} - \underbrace{f \cdot \mu_i}_{\text{back mutation}} + \underbrace{\sqrt{\frac{f \cdot (1-f)}{N}} \eta(t)}_{\text{genetic drift}} , \tag{11}$$

where we have assumed forward and backward mutations occur at the same locus-dependent rate μ_i. While these dynamics lead to tractable predictions, they will break down in adapting populations when the transfer rate begins to scale with the typical fitness variation. In these scenarios, evolutionary dynamics will sensitively depend on many interacting loci.

3 Rare beneficial sweeps

These interactions can again simplify when beneficial mutations are rare and sweep through the population one-by-one. In these settings, genetic diversity is limited and new mutations will typically arise on a clonal or approximately clonal background $\vec{g}$. The fate of a beneficial mutation will depend on its fitness effect s_b. This benefit is drawn from the distribution of fitness effects,

$$\mu(s|\vec{g}) = \sum_{\vec{g'}} M(\vec{g} \to \vec{g'}) \cdot \delta(s - X(\vec{g'}) + X(\vec{g})) , \tag{12}$$

which tabulates the rates and benefits of mutations at all sites. Provided selection is sufficiently strong $Ns_b \gg 1$, the new beneficial mutation will survive the stochastic fluctuations of genetic drift with a probability $\sim 2s_b$ and begin to grow deterministically. Its probability of fixing—and recombination's effect on the overall rate of fixation– will depend on whether the background is completely or approximately clonal. We review these two cases separately below.

3.1 Sweeps against a clonal background

In an otherwise clonal population, recombination can affect the fate of the sweeping mutation in two key ways. First, HGT events can transfer the mutation to the clonal background at a total rate $\sim r \cdot \overline{\ell} \cdot f(1-f)$. Simultaneously, transfers from the clonal population will overwrite the mutation at the same rate. Since these processes balance, recombination will have no net effect on the mutation's fate. This again leads to the standard single locus diffusion limit in Eq. (11),

$$\frac{df}{dt} = \underbrace{f \cdot (1-f) \cdot s_b}_{\text{selection}} + \underbrace{r \cdot \overline{\ell} \cdot f \cdot (1-f)}_{\text{mutation transferred}} - \underbrace{r \cdot \overline{\ell} \cdot f \cdot (1-f)}_{\text{mutation overwritten}} + \underbrace{\sqrt{\frac{f \cdot (1-f)}{N}}\eta(t)}_{\text{genetic drift}} , \tag{13}$$

where the successive sweeps assumption ensures forward and back mutation are negligible. Over multiple sweeps, these dynamics effectively resemble asexual adaptation. This implies recombination will also have no net effect on the rate of fixation or rate of adaptation. For the clonal sweeps regime to remain valid, each sweep will need to occur before another mutation is produced. This gives rise to a self-consistency condition that ensures the background remains completely clonal,

$$NU_b \ll \frac{\overline{s_b}}{\log(N\overline{s_b})} \ll 1, \tag{14}$$

where $\overline{s_b}$ is the average fitness effect of a beneficial mutation and $U_b \equiv \int_0^\infty \mu(s|\vec{g})\, ds$ is the beneficial mutation rate.

3.2 Initial breakdown of clonal sweeps

When multiple mutations arise in the population concurrently, recombination can begin to play a role in adaptation. In the simplest extension, all but one sweeping mutation will remain at small frequency at any given time. The majority of new beneficial mutations will be lost due to the stochastic fluctuations of genetic drift. These mutations can segregate at a small frequency between successive sweeps. Recombination can bring together mutations from the sweeping and drifting genotypes, creating exceptionally fit double mutants. This can occur provided sweeps are rare, but the background is not entirely clonal, which is satisfied in the limit,

$$\frac{\overline{s_b}}{\log(N\overline{s_b})} \ll NU_b \ll 1 \, . \tag{15}$$

The rate at which double mutant clones are generated will depend on the total size of the fluctuating class. Doomed beneficial mutations will arise at a total rate $\sim NU_b \cdot (1 - 2\overline{s_b})$ and reach a typical maximum frequency $\sim 1/2N\overline{s_b}$. This implies that the typical total frequency of these small frequency mutations will be $\sim U_b/\overline{s_b}$, which Eq. (15) ensures is a large number of individuals. Recombination events between the sweeping mutation and a doomed small frequency mutation will occur at a total rate $\sim Nr \cdot \overline{\ell} \cdot U_b/\overline{s_b}$. These recombined genotypes will survive drift and outcompete the sweeping mutation with a probability proportional to the fitness benefit of the recombinant: $\langle s_i + s_j + \epsilon_{i,j} \rangle_{i,j}$, which averages over the epistatic effects of both mutations. Provided this benefit remains comparable to $\overline{s_b}$, we obtain a self-consistency condition for successive sweeps to remain valid in the presence of recombination,

$$\frac{\overline{s_b}}{r \cdot \overline{\ell}} \gg NU_b \cdot \log(N\overline{s_b}) \, . \tag{16}$$

This predicted threshold has several interesting features. Firstly, along with Eq. (15), it suggests that recombination will not influence the rate of fixation in the successive mutations regime while $r \cdot \overline{\ell} \ll \overline{s_b}$. As we will show, the threshold rate in larger populations can be dramatically smaller. Secondly, Eq. (16) scales only linearly with both the mean fitness benefit and transfer length. This implies that the threshold rate is insensitive to other aspects of the underlying distribution of fitness effects or transfer lengths. We will also show that this simple scaling can break down in larger populations.

3.3 Deleterious mutations

Deleterious mutations can also reduce the rate of selective sweeps by reducing the number of viable backgrounds for a beneficial mutation to arise on. When this occurs, recombination can liberate beneficial mutations from poor backgrounds allowing them to sweep through the population. This process will become important when deleterious mutations reach an $\mathcal{O}(1)$ frequency on a timescale comparable with the typical sweep time. Deleterious mutations will occur at a total rate NU_d, where U_d is the per genome deleterious mutation rate. Provided the typical cost is large $N|s_d| \gg 1$, these mutations will reach a large frequency $f \sim 1/2$ with probability $\sim 1/N \cdot e^{-N|s_d|}$. This implies that recombination will have a negligible effect on the rate of fixation when,

$$\log\left(\frac{U_b}{U_d}\right) + \log\left(N\overline{s_b}\right) \ll N|s_d| \, . \tag{17}$$

Given empirical estimates of the fraction of mutations that are beneficial, the first term in Eq. (17) will often be no larger than $\mathcal{O}(1)$. As a result, this condition will typically be met in extremely small populations. When selection is comparatively weaker $N|s_d| \ll 1$, deleterious mutations will segregate quasi-neutrally and new deleterious mutations will reach an $\mathcal{O}(1)$ frequency with probability/ $\sim 1/N$. This suggests recombination will not influence the rate of fixation provided,

$$\log(N\overline{s_b}) \gg \log\left(\frac{U_d}{U_b}\right) \, . \tag{18}$$

This represents a weaker condition that is also often satisfied in moderate population sizes.

4 Clonal interference and recombination

In large populations $(NU_b \gg 1)$, multiple beneficial mutations can arise and attain a large frequency at the same time. This process is known as 'clonal interference' since the competition between variants can lead to many new mutations being wasted on poor genetic backgrounds. In principle, recombination can help alleviate clonal interference by recombining beneficial mutations onto a common genetic background. This process and the favorability of recombination events will depend on the potentially large number of epistatic interactions between competing variants in Eq. (10). To explore the basic role of recombination in breaking down clonal interference, we initially focus on a simple fitness landscape where there are no deleterious mutations and the fitness benefits s_i and mutation rates μ_i at each site remain the same for all genotypes. This allows us to express the distribution of fitness effects in terms of these rates and benefits,

$$\mu(s|\vec{g}) \equiv \sum_{i=1}^{L} \mu_i \cdot \delta(s - s_i) \, . \tag{19}$$

Consistent with the clonal interference regime we assume that a large number, but overall minority of sites are beneficial $1 \ll f_b \cdot L \ll L$.

Distribution of backgrounds The fate of a new mutation will strongly depend on the fitness of the genetic background on which it arises. A large body of work has shown that the distribution of backgrounds can often be simplified by considering a coarse-grained picture, which groups together individuals with the same overall fitness:

$$f(X,t) = \sum_{\vec{g}} f(\vec{g},t) \cdot \delta(X - X(\vec{g}))\,. \tag{20}$$

This allows Eq. (8) to be recast in terms of the dynamics of individual fitness classes,

$$\begin{aligned}\frac{\partial f(X)}{\partial t} = &\underbrace{\left[X - \overline{X}(t)\right] f(X)}_{\text{selection}} + \int ds\ \underbrace{\mu(s)\left[f(X-s) - f(X)\right]}_{\text{mutation}} + \underbrace{\int dX' \left[\delta(X - X') - f(X)\right] \sqrt{\frac{f(X')}{N}}\eta(X')}_{\text{genetic drift}} \\ + \quad & r \int dX' \int dX'' \sum_{\vec{g}' \in X'} \sum_{\vec{g}'' \in X''} \int d\ell\ \lambda(\ell)\ f(X', \vec{g}') \cdot f(X'', \vec{g}'') \sum_{i',i''=1}^{L-\ell} \times\,, \\ & \underbrace{\Big[\delta\left(X' - X + \textstyle\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i\right) + \delta\left(X' - X - \sum_{i'}^{i'+\ell} s'_i + \sum_{i''}^{i''+\ell} s''_i\right) - \\ & \delta(X - X') \cdot (1 - \delta(\textstyle\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i)) - \delta(X - X') \cdot (1 - \delta(\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i))\Big]}_{\text{recombination}}\end{aligned} \tag{21}$$

where $\overline{X}(t) = \int X f(X,t) dX$ is the mean fitness of the population, $\eta(X)$ is an uncorrelated Brownian noise term, and the single and double primed variables denote the acceptor and donor genotypes respectively. Previous work (33) (7) has shown that fluctuations in the fitness distribution are often small so that it can be described by a typical profile. In these scenarios, the mean fitness can be expanded in terms of its average behavior,

$$\overline{X}(t) \approx \langle \partial_t X \rangle|_{t=t_0} \cdot t + \frac{1}{2} \langle \partial_t^2 X \rangle\big|_{t=t_0} \cdot t^2 + \cdots\,. \tag{22}$$

The higher-order contributions to the mean fitness can typically be neglected such that the mean fitness can be approximated as increasing at an average rate $\langle \partial_t X \rangle \approx v$, referred to as the 'rate of adaption'. This allows us to express the typical profile of the fitness distribution in terms of the moving

'traveling wave' coordinate $x \equiv X - vt$,

$$0 = \underbrace{x \cdot f(x)}_{\text{selection}} - \underbrace{v \cdot \partial_x f(x)}_{\text{competition w/ population}} + \underbrace{\int ds\, \mu(s) \cdot [f(x-s) - f(x)]}_{\text{mutation}}$$

$$+ \quad r \int dx' \int dx'' \sum_{\vec{g}' \in x'} \sum_{\vec{g}'' \in x''} \int d\ell\, \lambda(\ell)\, f(x', \vec{g}') \cdot f(x'', \vec{g}'') \sum_{i',i''=1}^{L-\ell} \times$$

$$\Big[\delta\left(x' - x + \textstyle\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i\right) + \delta\left(x' - x - \textstyle\sum_{i'}^{i'+\ell} s'_i + \sum_{i''}^{i''+\ell} s''_i\right) -$$

$$\underbrace{\delta(x-x') \cdot (1 - \delta(\textstyle\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i)) - \delta(x-x') \cdot (1 - \delta(\textstyle\sum_{i'}^{i'+\ell} s'_i - \sum_{i''}^{i''+\ell} s''_i))]}_{\text{recombination}} \tag{23}$$

where we have neglected the non-linear drift terms in Eq. (4).

Fate of a new mutation A new mutation will arise on a background drawn from $f(x)$. As it competes for dominance in the population, recombination can influence the fate of the new mutation in several key ways. First, HGT events can transfer mutations from other genetic backgrounds to the new mutation's background. This process is similar in effect to producing linked mutations, since the focal mutation can hitchhike to fixation with the linked mutations it acquires. Unlike mutation, though, the fitness benefits of acquired mutations will not depend on $\mu(s)$, but rather the fitness benefits of mutations segregating at a large frequency that are most likely to be transferred. This same process can also present a disadvantage to a segregating mutation. Transfer events can overwrite mutations on the new mutations background, decreasing its fitness. Notably, overwriting actually becomes more prominent on fitter backgrounds which tend to have more variants than a randomly sampled donor background.

The new mutation can also take over by jumping to an entirely different genetic background itself. The contributions from this process will strongly depend on the distribution of acceptor backgrounds which are randomly drawn from $f(x)$. Although most transfers will land the mutation on a less fit or roughly equivalent genetic background, rare transfers can lead to atypically fit backgrounds. This process will depend on the genetic position of the focal mutation relative to other mutation's on its original background. This is because the focal mutation can be transferred along with any colocalized mutations conferring a larger benefit to the acceptor genome.

This competition is summarized by the non-extinction probability $w(g_j|x, \vec{g})$, which corresponds to the probability that a new mutation at locus j, on genetic background $\vec{g}$, with fitness x takes over in the population. Previous work has shown that the fate of the new mutation will be determined at small frequencies (32) (34). This allows us to linearize the drift term and write the down a branching process

recursion for $w(g_j|x,\vec{g})$,

$$0 = \underbrace{x\cdot w(g_j|x,\vec{g})}_{\text{selection}} - \underbrace{v\cdot\partial_x w(g_j|x,\vec{g})}_{\text{competition w/ population}} + \underbrace{\sum_{i=1}^{f_b\cdot L}\mu_i\cdot[w(g_j|x+s_i,\vec{g}+g_i)-w(g_j|x,\vec{g})]}_{\text{mutation}} - \underbrace{\frac{w(g_j|x,\vec{g})^2}{2}}_{\text{genetic drift}}$$
$$+ \quad r\int dx'\sum_{\vec{g}'\in x'}\int d\ell\;\lambda(\ell)\;f(x',\vec{g}')\cdot w(g_j|x'',\vec{g}'')\sum_{i,i'=j-\ell}^{j}\times$$
$$\underbrace{\left[\delta\left(x''-x'-\textstyle\sum_i^{i+\ell}s_i+\sum_{i'}^{i'+\ell}s_i'\right)-\tfrac{1}{2}\right]\cdot\left[\delta\left(\vec{g}''-\vec{g}'-\textstyle\sum_i^{i+\ell}g_i+\sum_{i'}^{i'+\ell}g_i'\right)-\tfrac{1}{2}\right]}_{\text{mutation(s) transferred}}$$
$$+ \quad r\int dx'\sum_{\vec{g}'\in x'}\int d\ell\;\lambda(\ell)\;f(x',\vec{g}')\cdot w(g_j|x'',\vec{g}'')\sum_{i,i'\notin(j-\ell,j)}\times$$
$$\underbrace{\left[\delta\left(x''-x+\textstyle\sum_i^{i+\ell}s_i-\sum_{i'}^{i'+\ell}s_i'\right)-\tfrac{1}{2}\right]\cdot\left[\delta\left(\vec{g}''-\vec{g}+\textstyle\sum_i^{i+\ell}g_i-\sum_{i'}^{i'+\ell}g_i'\right)-\tfrac{1}{2}\right]}_{\text{mutation(s) acquired/overwritten}} \tag{24}$$

With this framework in place, the rate of adaptation can be self-consistently calculated by averaging over the backgrounds a new mutation can arise on and the fitness benefit it confers,

$$v\approx\sum_{i=1}^{L}N\mu_i\int d\vec{g}\int dx\;f(x)\cdot w(x+s_i|x,\vec{g}+g_i)\cdot s_i\;. \tag{25}$$

While Eq. (23) and Eq. (24) are simple to write down, they are extremely difficult to solve in practice. This is because the recombination kernel depends on the exact genotype of all individuals in the population. To make progress, a large body of work has focused on two asymptotic limits where the genotype dependence dramatically simplifies (7)(34)(33). We refer to the first of these limits as the 'rapid recombination' regime since it assumes recombination occurs rapidly enough to ensure that all established mutations are in complete linkage equilibrium. This implies that a mutation's chance of belonging to a given genotype is simply proportional to its frequency in the population. The second oft-studied limit is asexual adaptation where there is no recombination. In the 'asexual limit' there is strong linkage disequilibrium since mutations typically reach a large frequency on a single genetic background. These limits are important to contextualize our contribution, which focuses on the intermediate regime where clonal interference is pervasive and recombination acts too slowly to resolve linkage disequilibrium. Below, we briefly review the key results of these previous approaches before showing how they can be extended.

5 Fast recombination limit

Distribution of fitness In the rapid recombination limit, clonal interference influences the fate of a mutation while it competes to establish in the population. At these small frequencies, the contributions from acquiring mutations will depend on the number and benefits of mutations on a random donor and acceptor genotype. To simplify this process, previous work assumed that the mean transfer length is sufficiently small that at most one mutation can be transferred in a single event. This allowed them to

neglect linkage during transfer events and rewrite the contributions from acquiring and overwriting mutations in Eq. (23) in terms of the frequencies of individual mutations,

$$r\cdot\overline{\ell}\cdot\sum_{i=1}^{L}\int dx' f(x')\cdot[\underbrace{h(g_i|x',t)\cdot(1-h(g_i|x-s_i,t))\cdot f(x-s_i)}_{\text{mutation acquired}}+\underbrace{(1-h(g_i|x',t))\cdot h(g_i|x+s_i,t)\cdot f(x+s_i)}_{\text{mutation overwritten}}], \tag{26}$$

where $f(x)$ is the frequency of the fitness class with relative fitness x and $h(g_i|x)$ is the frequency of mutation g_i within fitness class x. To further simplify, previous work has assumed that all mutations confer the same benefit s_b and occur at the same total rate U_b. This ensures that all transfer events also confer the same benefit. For the relevant values of x, Eq. (26) can then be approximated by expanding around s_b small,

$$\begin{aligned} &r\cdot\overline{\ell}\,\frac{\partial}{\partial x}s_b\,\left\langle\sum_{i=1}^{L}\int dx' f(x')\cdot f(x)\left[-h(g_i|x',t)\cdot(1-h(g_i|x,t))+(1-h(g_i|x',t))\cdot h(g_i|x,t)\right]\right\rangle_t \\ &\quad+r\cdot\overline{\ell}\,\frac{\partial^2}{\partial x^2}\frac{s_b}{2}\,\left\langle\left(\sum_{i=1}^{L}\int dx' f(x')\cdot f(x)\left[-h(g_i|x',t)\cdot(1-h(g_i|x,t))+(1-h(g_i|x',t))\cdot h(g_i|x,t)\right]\right)^2\right\rangle_t+\dots\,. \end{aligned} \tag{27}$$

Crucially in linkage equilibrium, the frequency of a mutation does not depend on its fitness background, so $h(g_i|x',t)=h(g_i|x,t)$ and the contribution from recombination entirely disappears. This ensures that the dominant balance in Eq. (23) simply arises from the first two-terms which has an exactly Gaussian solution,

$$f(x)\approx\frac{e^{\frac{-x^2}{2v}}}{\sqrt{2\pi v}}\,. \tag{28}$$

Non-extinction probability of a new mutation By enforcing the same assumptions, the recombination kernel in Eq. (24) simplifies to,

$$\begin{aligned} &r\cdot\overline{\ell}\,\frac{\partial}{\partial x}s_b\,\left\langle\sum_{i=1}^{L}\int dx' f(x')\cdot w(g_j|x)\left[\underbrace{-h(g_i|x',t)\cdot(1-h(g_i|x,\vec{g},t))}_{\text{mutation acquired}}+\underbrace{(1-h(g_i|x',t))\cdot h(g_i|x,\vec{g},t)}_{\text{mutation overwritten}}\right]\right\rangle_t \\ &\quad+r\cdot\overline{\ell}\,\frac{\partial^2}{\partial x^2}\frac{s_b}{2}\,\left\langle\left(\sum_{i=1}^{L}\int dx' f(x')\cdot w(g_j|x)\left[\underbrace{-h(g_i|x',t)\cdot(1-h(g_i|x,\vec{g},t))}_{\text{mutation acquired}}+\underbrace{(1-h(g_i|x',t))\cdot h(g_i|x,\vec{g},t)}_{\text{mutation overwritten}}\right]\right)^2\right\rangle_t \\ &\quad+\underbrace{r\cdot\tilde{\ell}\int f(x')\cdot w(g_j|x'+s_b)dx'}_{\text{mutation transferred}}, \end{aligned} \tag{29}$$

where we have included contributions from the mutation transferring to a new background. Since we are now focused on the frequency of g_i on a specific genotype, the contributions from acquiring and overwriting mutations no longer disappears. In fact, it is straightforward to show that a second-order

expansion yields the simplified differential equation for $w(g_j|x)$,

$$0 = \underbrace{x \cdot w(g_j|x)}_{\text{selection}} - \underbrace{v \cdot \partial_x w(g_j|x)}_{\text{competition w/ population}} + \underbrace{U_b \cdot [w(g_j|x+s_b) - w(g_j|x)]}_{\text{mutation}} - \underbrace{\frac{w(g_j|x)^2}{2}}_{\text{genetic drift}} \tag{30}$$

$$+ \underbrace{r \cdot \bar{\ell} \left[x \frac{\partial}{\partial x} + v \frac{\partial^2}{\partial x^2} \right] w(g_j|x) + r \cdot \tilde{\ell} \int f(x') \cdot w(g_j|x'+s_b) dx'}_{\text{recombination}} , \tag{31}$$

where we have assumed linkage equilibrium. This kernel highlights a crucial feature of the evolutionary dynamics in the rapid recombination regime. We can think of recombination as a diffusive process spreading a mutation across many backgrounds with a similar fitness. Transfer events will typically overwrite mutations on exceptionally fit backgrounds, pushing them towards the mean fitness. However, HGT can also more rarely lead to acquisition events on these backgrounds that allow a mutation to 'surf' towards higher fitnesses. This highlights that there is no clonal structure in the rapid recombination regime. A mutation will typically spread to many backgrounds before it establishes and takes over.

Previous work has solved Eq. (30) in the rapid recombination limit where contributions from linked mutation can be neglected. This leads to a rate of adaptation that scales as,

$$v \approx r \cdot \bar{\ell} \cdot \sqrt{2 \log N U_b} \, . \tag{32}$$

These results and the rapid recombination regime are valid provided,

$$s_b \log(N U_b) \ll r \cdot \bar{\ell} \ll s_b \sqrt{\frac{N U_b}{\log(N U_b)}} \, , \tag{33}$$

where larger recombination rates coincide with the the independent sites limit. This represents a range of extremely large recombination rates. In practice, many microbial populations may exchange genetic material at far lower rates where the linkage equilibrium approximation breaks down. Provided recombination is sufficiently rare, we can turn to the alternative oft-studied asexual limit.

6 Asexual limit

In the absence of recombination, the dynamics in Eq. (23) and Eq. (24) do not have any explicit genotype dependence. This allows us to solve straightforwardly for the distribution of fitness. Previous work has shown that $f(x)$ will have a maximum value x_{ca} corresponding to the relative fitness of the fittest individual. Below this threshold, the shape of $f(x)$ will be determined by the adaptation of the population, selection against the mean fitness and the flux of mutation events,

$$0 = \underbrace{x \cdot f(x)}_{\text{selection}} - \underbrace{v_a \cdot \partial_x f(x)}_{\text{competition w/ population}} + \underbrace{\int \mu(s) \left[f(x-s) - f(x) \right] ds}_{\text{mutation}} \, . \tag{34}$$

In our regime of interest, the contributions from mutation in Eq. (34) will be negligible such that the distribution of fitness will be approximately Gaussian below x_{ca},

$$f(x) \approx \begin{cases} 0 & \text{if } x > x_{ca} \\ \frac{e^{\frac{-x^2}{2v_a}}}{\sqrt{2\pi v_a}} & \text{if } x < x_{ca} \, . \end{cases} \tag{35}$$

Previous work has also shown that the fitness benefits of successful mutations will typically be concentrated around a characteristic value $\tilde{s_b}$, which occur at a total rate $\tilde{U_b}$. The solution in Eq. (35) will remain valid provided $x_{ca} - \tilde{s_b} \gg \sqrt{v_a}$ and $\tilde{s_b} \gg \sqrt{v_a}$, which constitutes the formal definition of the regime that we consider in this work. The first condition ($x_{ca} - \tilde{s_b} \gg \sqrt{v_a}$) implies that the parents of the most fit individuals are substantially more fit than the majority of the population (i.e. clonal interference is common). The second condition ($\tilde{s_b} \gg \sqrt{v_a}$) implies that the majority of the population is concentrated near the mean fitness. This is also known as the "moderate speeds" regime, since the second condition can be rewritten as $v_a \ll \tilde{s_b}^2$. These assumptions are equivalent to assuming,

$$\log(\tilde{s_b}/\tilde{U_b}) \ll 2 \cdot \log(N\tilde{s_b}) \ll \log^2(\tilde{s_b}/\tilde{U_b}) \,, \tag{36}$$

which require that $N\tilde{s_b} \gg 1$ and $\tilde{s_b} \gg \tilde{U_b}$.

Branching process solution for the fixation probability A new mutation will arise on a background drawn from $f(x)$. The fate of this mutation can be described by the conditional fixation probability $w_a(g_j|x)$. Provided the fate of the mutation is determined at small frequencies, $w_a(g_j|x)$ will obey the branching process recursion,

$$0 \approx \underbrace{x \cdot w_a(g_j|x)}_{\text{selection}} + \underbrace{\int \mu(s) \left[w_a(g_j|x+s) - w_a(g_j|x)\right] ds}_{\text{mutation produced}} - \underbrace{v_a \cdot \partial_x w_a(g_j|x)}_{\text{adaptation of the population}} - \underbrace{\frac{1}{2} \cdot w_a(g_j|x)^2}_{\text{genetic drift}} \,, \tag{37}$$

which represents a balance between the (i) the growth of a clone due to selection, (ii) the adaptation of the background population, (iii) the stochastic effects of genetic drift, and (iv) the production of further mutations. Previous work has shown that for large values of fitness $x \gtrsim x_{ca} - \tilde{s_b}$ the solution to $w_a(x)$ is well approximated by the 'shoulder solution' where the contributions from mutation can be neglected,

$$w_a(g_j|x) \approx \begin{cases} 2x & \text{if } x > x_{ca} \\ 2x_{ca} e^{\frac{x^2 - x_{ca}^2}{2v_a}} & \text{if } x_c - \tilde{s_b} < x < x_{ca} \,. \end{cases} \tag{38}$$

Below x_{ca}, contributions from the nonlinear term are negligible and a recursive solution of $w_a(x)$ can be obtained from an integral transform of Eq. (37),

$$w_a(g_j|x) = \int_0^\infty dt \int \mu(s) \cdot \underbrace{e^{xt - \frac{v_a t^2}{2}}}_{\text{size of clone}} \cdot \underbrace{w_a(x - v_a t + s)\, ds}_{\text{fixation prob of offspring}} \,. \tag{39}$$

We refer to this solution as the 'clone expansion' since it summarizes the fixation probability of a clone by the fixation probability of its offspring. It can also be equivalently expressed in terms of the time-dependent relative fitness of the founding lineage $y = x - v_a t$,

$$w_a(g_j|x) = \frac{e^{\frac{x^2}{2v_a}}}{v_a} \cdot \int_0^x dy \int \mu(s) \cdot e^{-\frac{y^2}{2v_a}} \cdot w_a(y+s)\, ds \,. \tag{40}$$

The clone expansion and distribution of backgrounds allow us to obtain physical intuition for how a successful mutation will take over. It will typically arise near the nose of the fitness distribution and begin to grow exponentially while producing additional linked mutations. The fitness benefits of these mutations $\tilde{s_b}$ will correspond to the value of s where the integrand is maximized in Eq. (40). The mutation's offspring will tend to land near the interference threshold x_c, ensuring the mutation remains in the nose while it takes over.

Location of the interference threshold A condition for the location of the interference threshold can be obtained by requiring that the shoulder solution and clone expansion yield the same solution in the region $x_{ca} - \tilde{s_b} \lesssim x \lesssim x_{ca}$,

$$1 = \frac{1}{v_a} \cdot \int_0^{x_c} dy \int \mu(s) \cdot e^{-\frac{y^2}{2v_a}} \cdot w_a(y+s)\, ds \,. \tag{41}$$

This yields the relation,

$$1 \approx \left(\frac{\tilde{U_b}}{\tilde{s_b}}\right) \exp\left[\frac{x_{ca}\tilde{s_b}}{v_a} - \frac{\tilde{s_b}^2}{2v_a}\right] \left(\frac{x_{ca}}{x_{ca} - \tilde{s_b}}\right) \,, \tag{42}$$

which can be used to solve iteratively for x_{ca} in terms of v_a.

Rate of adaptation The rate of adaptation obeys another self-consistency condition,

$$v_a = NU_b \int dx \int \mu(s) \cdot s \cdot f(x) \cdot w_a(x+s) ds \,, \tag{43}$$

which simply integrates the fitness gain of successful mutations. This condition can be evaluated to obtain,

$$1 \approx \frac{2N x_{ca} \tilde{s_b} e^{-\frac{x_{ca}^2}{2v_a}}}{\sqrt{2\pi v_a}} \,. \tag{44}$$

Iterative solutions to these equations in our regime of interest, yield the approximate scaling,

$$v_a \approx \frac{2\tilde{s_b}^2 \log(N\sqrt{\tilde{s_b} \cdot \tilde{U_b}})}{\log^2(\tilde{s_b}/\tilde{U_b})} \,, \qquad x_{ca} \approx \frac{2\tilde{s_b} \log(N\tilde{s_b})}{\log(\tilde{s_b}/\tilde{U_b})} - 1/2 \,. \tag{45}$$

Rate of fixation With this formalism, the total substitution rate of mutations can can also be calculated straightforwardly,

$$\theta_a = NU_b \cdot \int dx \int \mu(s) \cdot f(x) \cdot w_a(x+s) ds \,. \tag{46}$$

Substituting the solution for $w_a(x)$, the rate of fixation scales as

$$\theta_a \sim U_b \cdot e^{\frac{x_{ca}\tilde{s_b}}{v_a} - \frac{\tilde{s_b}^2}{2v_a}} \,. \tag{47}$$

7 Quasi-asexual theory

To begin to consider the intermediate regime between the asexual and rapid-recombination limits, we initially focus on the simplest extension of asexual dynamics. In this regime, recombination will act too slowly to break apart the typical new mutation from its original background before it takes over or is purged from the population. We refer to this regime as the 'quasi-asexual' limit since recombination will remain sufficiently rare such that θ_a will represent the dominant contribution to the overall substitution rate. This implies that contributions from HGT will remain negligible in Eq. (34) and the fitness distribution will retain its typical asexual shape. As a result, we will be able to consider the fate of a rare recombined mutation that arises on $f(x)$ and competes against a predominantly asexual background population and rate of adaptation. As we allow for increasing amounts of recombination, we can ask: at what recombination rates do recombined mutations take over on a par with those on their original genetic background?

7.1 Contributions to the branching process recursion for the fixation probability

Contributions from acquiring mutations Consistent with (7), we initially focus on the simplest distribution of fitness effects where all mutations confer the same fitness benefit s_b and occur at the same per base pair rate μ,

$$\mu(s) \equiv U_b \cdot \delta(s - s_b) \,, \tag{48}$$

where $U_b \equiv \mu \cdot L \cdot f_b$. This allows us to assume that the fitness benefits of produced and transferred mutations are equal to s_b. The rate of acquisition events will depend on the genetic diversity and descent of genotypes within the population. This is because an acceptor genotype cannot acquire mutations that it already possesses. The overall rate at which the focal mutation's background acquires mutations will depend on the number of mutations unique to a random donor drawn from $f(x)$. As we will discuss, successful transfers will generally involve an acceptor with greater fitness than a donor. Provided recombination is sufficiently rare, previous work has shown we can approximate the distribution of k as a simple function of the acceptor and donor's fitness x and x',

$$\Pr\left(k, x, x'\right) \approx \begin{cases} \frac{1}{\tilde{k}} \cdot \left(1 - \frac{1}{\tilde{k}}\right)^{k - \frac{x}{s_b} + \frac{x'}{s_b}} & \text{if } k \cdot s_b > x' - x \,, \\ 0 \text{ otherwise} \,. \end{cases} \tag{49}$$

where the expected value $\tilde{k}$ is equal to the typical number of mutations separating the fittest individuals from those at the mean fitness x_c/s_b. For the empirically relevant scenarios we consider, this typical number will remain small $(1 \lesssim \tilde{k} \lesssim 10)$, so that the typical transfer event involves at most a single mutation: $\tilde{k} \cdot \bar{\ell} \ll L$. However, even while acquisition events involving multiple mutations are rare, they can provide a multiplicative fitness benefit to the acceptor genome when they do occur. The rate of acquiring multiple mutations will depend on the genetic distances $\delta\ell$ between adjacent mutations on a random donor genotype. This distribution follows straightforwardly from the random uniform distribution of beneficial loci,

$$\Pr(\delta\ell) \equiv \frac{k}{L} \cdot e^{-\frac{\delta\ell}{L/k}} \,. \tag{50}$$

Most transfer lengths will be smaller than this typical spacing, in which case Eq. (50) simplifies to a uniform distribution, allowing us to approximate the rate of acquiring n mutations from a random donor $\tilde{R}_{n\leftarrow}$ as,

$$\tilde{R}_{n\leftarrow} \approx r \cdot L \int x' dx' \sum_k \Pr(k, x, x') \int \lambda(\ell) \cdot \binom{k}{n} \cdot \left(\frac{\ell}{L}\right)^n \cdot \left(1 - \frac{\ell}{L}\right)^{k-n} d\ell, \tag{51}$$

where the left arrow denotes mutations incoming to the focal mutation's genome. As we will show, the typical successful transfer event will involve an acceptor with larger fitness than a donor. This allows us to neglect the x' donor dependence in $\Pr(k)$ and approximate,

$$\tilde{R}_{n\leftarrow} \sim r \cdot L \cdot \int \lambda(\ell) \cdot \left(\frac{\tilde{k} \cdot \ell}{L}\right)^n d\ell \,, \tag{52}$$

where we have expanded Eq. (51) around $\ell \ll L$. This implies the role of recombination will depend on the balance between the rarity at which mutations are colocalized $\sim \int \lambda(\ell) \cdot (\ell/L)^n$ and the benefit they provide $\sim n \cdot s_b$.

Contributions from overwriting mutations We can use a similar logic to understand the rate at which mutation's are overwritten. The process of acquiring mutations will be equivalent to overwriting mutations, except the x and x' in Eq. (49) will be flipped since the focal background is accepting an empty piece of genetic material. Since the difference in donor and acceptor fitness will be maximally $x_c \sim \tilde{k} \cdot s_b$, the probability of overwriting mutations will be at most a factor of two larger than the rate in Eq. (52). This allows us to straightforwardly approximate the rate at which focal mutations are overwritten on the focal mutation's background with Eq. (52).

Contributions from transferring to a new genetic background alone The focal mutation can also take over by jumping to an entirely different genetic background itself. The contributions from this process will strongly depend on the distribution of acceptor backgrounds which are also randomly drawn from $f(x)$. Although most transfers will land the mutation on a less fit or roughly equivalent genetic background, rare transfers can lead to atypically fit backgrounds. This process will also depend on the genetic position of the focal mutation relative to other mutation's on its original background. For example, if a new mutation arises near another preexisting mutation, the two mutations' fate can become linked if they are likely to be transferred together. Since the dynamics are predominantly asexual, the genetic distance separating a new mutation from its nearest neighbor $\delta\ell_2$ will also be randomly sampled from the exponential distribution in Eq. (50). This ensures that for typical values of $\delta\ell_2$, the focal mutation will be overwhelmingly likely to be transferred alone at a total rate,

$$\tilde{R}_{1\rightarrow} = r \cdot \int_0^{\delta\ell_2} \ell \cdot \lambda(\ell)\, d\ell + r \cdot \int_{\delta\ell_2}^{L} \delta\ell_2 \cdot \lambda(\ell)\, d\ell\,, \tag{53}$$

where the right arrow denotes the focal mutation being transferred away from its present genotype.

Contributions from transferring to a new genetic background with colocalized mutations More rarely, the focal mutation will arise near a preexisting mutation, in which case the colocalized pair can be transferred together at a total rate,

$$\tilde{R}_{2\rightarrow} = r \cdot \int_{\delta\ell_2}^{L} \lambda(\ell) \cdot (\ell - \delta\ell_2)\, d\ell\,. \tag{54}$$

More generally, the focal mutation can also be transferred with n multiple other mutations on its background. The total probability that this occurs will depend on the distance separating the two outermost mutations in the group. We refer to this distance as $\delta\ell_n$, whose distribution will also follow straightforwardly from the random uniform distribution of beneficial loci,

$$\Pr(\delta\ell_n) \propto \frac{n(n-1)}{L} \cdot \left(\frac{\delta\ell_n}{L}\right)^{n-2} \cdot \left(1 - \frac{\delta\ell_n}{L}\right)^{n-2}\,. \tag{55}$$

The total rate that the focal mutation is transferred with $n-1$ other colocalized mutations will scale as,

$$\tilde{R}_{n\rightarrow} = r \cdot \int_{\delta\ell_n} \lambda(\ell) \cdot (\ell - \delta\ell_n)\, d\ell\,. \tag{56}$$

While colocalization events are rare, they can also provide a multiplicative fitness advantage to the acceptor genome. From the perspective of a new mutation, this represents an important difference

from acquiring multiple mutations: the probability of becoming colocalized scales linearly with the transferred length $\int \lambda(\ell) \cdot \ell/L$, after which the mutation can recoup a $n \cdot s_b$ with each subsequent transfer. As we will show the contributions from larger values of $\tilde{R}_{n\rightarrow}$ will become important for increasing amounts of clonal interference. For this reason, we limit our present focus to colocalized pairs and later show how to extend our approach to consider larger values of n. This approximation will be valid in the asymptotic limit $\overline{\ell}/L \to 0$ or $f_b \to 0$; however, we will show that it holds for much larger typical transfer lengths.

7.2 Branching process recursion with recombination

These assumptions allow us to again simplify the genotype dependence in Eq. (24). Specifically, the conditional fixation probability of a focal mutation that takes over via recombination now depends only on the genetic distance to its nearest neighbor $w_r(g_j|x, \delta\ell_2)$. To simplify notation, we refer to the quasi-asexual fixation probability,

$$w_r(x, \delta\ell_2) \equiv w_r(g_j|x, \delta\ell_2) \; . \tag{57}$$

This conditional fixation probability will satisfy a generalized version of the branching process recursion in Eq. (1),

$$\begin{aligned}
0 \approx & \underbrace{x \cdot w_r(x,\delta\ell_2)}_{\text{selection}} - \underbrace{v \cdot \partial_x w_r(x,\delta\ell_2)}_{\text{adaptation of the population}} - \underbrace{\frac{1}{2} \cdot w_r(x,\delta\ell_2)^2}_{\text{genetic drift}} + \underbrace{U_b \int_{\delta\ell_2}^{L} \Pr(\delta\ell) \cdot [w_r(x+s_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{distant mutation produced}} \\
& + \underbrace{U_b \int_{0}^{\delta\ell_2} \Pr(\delta\ell) \cdot [w_r(x+s_b,\delta\ell) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{more colocalized mutation produced}} + \underbrace{\sum_n \frac{\tilde{R}_{n\leftarrow}}{2} \int_{\delta l_2}^{L} \Pr(\delta\ell) \cdot [w_r(x+ns_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{distant mutation(s) acquired}} \\
& + \underbrace{\frac{\tilde{R}_{1\leftarrow}}{2} \int_0^{\ell} \frac{dz}{\ell} \int_z^{\delta l_2} \Pr(\delta\ell) \cdot [w_r(x+ns_b,\delta\ell) - w_r(x,\delta\ell_2)]\, d\delta\ell}_{\text{more colocalized mutation acquired}} + \underbrace{\sum_n \frac{\tilde{R}_{n\leftarrow}}{2} [w_r(x-ns_b,\delta\ell_2) - w_r(x,\delta\ell_2)]}_{\text{mutation(s) overwritten}} \\
& + \underbrace{\frac{\tilde{R}_{1\rightarrow}}{2} \int_0^{\ell} \frac{dz}{\ell} \int_z \Pr(\delta\ell) \int f(x) \cdot [w_r(x+s_b,\delta\ell) - w_r(x,\delta\ell)]\, dx}_{\text{focal mutation transferred alone}} + \underbrace{\frac{\tilde{R}_{2\rightarrow}}{2} \int f(x)\, [w_r(x+2s_b,\delta\ell_2) - w_r(x,\delta\ell_2)]\, dx}_{\text{focal mutation transferred as pair}} \; ,
\end{aligned} \tag{58}$$

which represents a balance between the (i) the growth of a clone due to selection, (ii) the adaptation of the background population, (iii) the stochastic effects of genetic drift, (iv) the production of an additional linked mutation, (v) the acquisition of one or more mutations, (vi) the overwriting of existing mutations via HGT, (vii) the transfer of the focal mutation alone to a new background, and (viii) the transfer of the focal mutation and a colocalized mutation to a new background. We have also included the possibility that the focal mutation's distance to its nearest neighbor can change from producing or acquiring an even more proximal mutation or by switching genetic background. These pairs founded by recombination occur at a slightly smaller rate because of the probability one of the two mutations gets overwritten during the transfer event.

The total substitution rate can then be obtained by averaging over the nearest neighbor distances and

initial fitness backgrounds that new mutations can arise on,

$$\theta_r = NU_b \int \Pr(\delta\ell_2)\, d\delta\ell_2 \int f(x) \cdot w_r(x + s_b, \delta\ell_2)\, dx\ . \tag{59}$$

This provides a natural scale to measure the effect of recombination. When $\theta_r \ll \theta_a$, most mutations will take over on their original background. As θ_r becomes larger than θ_a, most mutations will take over on a background different from their original one. When this occurs, recombination will transition to become the dominant force driving mutations to fix and our assumptions about the shape of the distribution of fitness and genetic diversity will break down.

7.3 Solution for the branching process recursion

To obtain a solution for $w_r(x, \delta\ell)$ we apply an ansatz similar to that used in the asexual case. We will assume that the mutation and recombination contributions can be neglected above a critical threshold $x_{cr} \equiv x_{cr}(\delta\ell_2)$ that depends on the value $\delta\ell_2$. This yields the familiar shoulder solution,

$$w_r(x, \delta\ell_2) \approx \begin{cases} 2x & \text{if } x > x_{cr} \\ 2x_{cr} e^{\frac{x^2 - x_{cr}^2}{2v}} & \text{if } x_{\text{min,r}} < x < x_{cr}\ . \end{cases} \tag{60}$$

Below x_{cr}, we can obtain a recursive solution analogous to the clone expansion of $w_a(x)$ via an integral transform of Eq. (58),

$$\begin{aligned}
w_r(x, \delta\ell_2) \approx\ & \underbrace{\frac{U_b \cdot e^{\frac{x^2}{2v}}}{v} \int_{\delta\ell_2}^{L} \Pr(\delta\ell)\, d\delta l \int^{x} e^{\frac{-y^2}{2v}} \cdot w_r(y + s_b, \delta\ell_2)\, dy}_{\text{distant mutation produced}} \\
& + \underbrace{\frac{U_b \cdot e^{\frac{x^2}{2v}}}{v} \int_{0}^{\delta\ell_2} \Pr(\delta\ell)\, d\delta l \int^{x} e^{\frac{-y^2}{2v}} \cdot w_r(y + s_b, \delta\ell)\, dy}_{\text{more colocalized mutation produced}} \\
& + \underbrace{\sum_n \frac{\tilde{R}_{n\leftarrow} \cdot e^{\frac{x^2}{2v}}}{2v} \int_{\delta l_2}^{L} \Pr(\delta\ell)\, d\delta\ell \int^{x} e^{\frac{-y^2}{2v}} \cdot w_r(y + n s_b, \delta\ell_2)\, dy}_{\text{distant mutation(s) acquired}} \\
& + \underbrace{\frac{\tilde{R}_{1\leftarrow} \cdot e^{\frac{x^2}{2v}}}{2v} \int_{0}^{\ell} \frac{dz}{\ell} \int_{z}^{\delta l_2} \Pr(\delta\ell)\, d\delta\ell \int^{x} e^{\frac{-y^2}{2v}} \cdot w_r(y + s_b, \delta\ell)\, dy}_{\text{more colocalized mutation acquired}} \\
& + \underbrace{\frac{\tilde{R}_{1\rightarrow} \cdot e^{\frac{x^2}{2v}}}{2v} \int_{0}^{\ell} \frac{dz}{\ell} \int_{z} \Pr(\delta\ell)\, d\delta\ell \int^{x} e^{\frac{-z^2}{2v}}\, dz \int^{x_c} \frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}} \cdot w_r(y + s_b, \delta\ell) dy}_{\text{focal mutation transferred alone}} \\
& + \underbrace{\frac{\tilde{R}_{2\rightarrow} \cdot e^{\frac{x^2}{2v_a}}}{2v} \int^{x} e^{\frac{-z^2}{2v}}\, dz \int^{x_c} \frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}} \cdot w_r(y + 2s_b, \delta\ell_2)\, dy}_{\text{focal mutation transferred as pair}}\ .
\end{aligned} \tag{61}$$

The location of the interference threshold can be obtained by requiring that the shoulder solution and clone expansion overlap in a boundary region $x_{\text{min,r}} \lesssim x \lesssim x_{cr}$.

8 Solution for a simple model of distribution of transfer lengths

To understand this crossover, we start by considering a simple model of the distribution of transfer lengths, where all transferred segments have the same length $\tilde{\ell}$.

$$\lambda(\ell) \equiv \delta(\ell - \tilde{\ell}) \; . \tag{62}$$

This implies that the rate of acquiring multiple (n) mutations will simply scale as $(\tilde{\ell}/L)^n$,

$$\tilde{R}_{n\leftarrow} \approx r \cdot \tilde{\ell} \cdot \tilde{k} \cdot \left(\frac{\tilde{k} \cdot \tilde{\ell}}{L} \right)^{n-1} \; . \tag{63}$$

It also allows us to clearly define colocalized pairs as having a nearest neighbor distance smaller than the transfer length $\delta\ell_2 \lesssim \tilde{\ell}$ since both mutations will typically be transferred together at a total rate,

$$\tilde{R}_{2\rightarrow} = r \cdot (\tilde{\ell} - \delta\ell_2) \; . \tag{64}$$

On the other hand, when the initial separation exceeds the transfer length $\delta\ell_2 \geq \tilde{\ell}$, the new mutation will always be transferred alone at a total rate,

$$\tilde{R}_{1\rightarrow} = r \cdot \tilde{\ell} \; . \tag{65}$$

These simplifications allow us to define separate branching processes $w_{r<}(x, \delta\ell_2)$ and $w_{r>}(x, \delta\ell_2)$ corresponding to $\delta\ell_2 < \tilde{\ell}$ and $\delta\ell_2 > \tilde{\ell}$ respectively. The substitution rate in Eq. (59) can be separated by these contributions,

$$\theta_r \approx \underbrace{NU_b \int_0^{\tilde{\ell}} \Pr(\delta\ell_2) \, d\delta\ell_2 \int f(x) \cdot w_{r<}(x + s_b, \delta\ell_2) \, dx}_{\theta_{r>} \text{ initially colocalized}} + \underbrace{NU_b \cdot \int f(x) \cdot w_{r>}(x + s_b, \delta\ell_2) \, dx}_{\theta_{r<} \text{ not initially colocalized}} \; , \tag{66}$$

where we have enforced our assumption that $\tilde{k} \cdot \tilde{\ell} \ll L$. To understand which term represents the dominant contribution, we consider these branching processes separately below.

8.1 Mutations that are not initially colocalized

Provided a new mutation is initially colocalized, we can simplify the clone expansion solution in Eq. (58),

$$
\begin{aligned}
w_{r>}(x,\delta\ell_2) \approx & \underbrace{\frac{U_b\cdot e^{\frac{x^2}{2v}}}{v}\int^{x} e^{\frac{-y^2}{2v}}\cdot w_{r>}(y+s_b,\delta\ell_2)\,dy}_{\text{distant mutation produced}} + \underbrace{\frac{U_b\cdot e^{\frac{x^2}{2v}}}{v}\int_0^{\tilde{\ell}}\Pr(\delta\ell)\,d\delta\ell\int^{x} e^{\frac{-y^2}{2v}}\cdot w_{r<}(y+s_b,\delta\ell)\,dy}_{\text{colocalized mutation produced}} \\
& + \underbrace{\sum_n \frac{r\cdot\tilde{\ell}\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v}\cdot\left(\frac{\tilde{\ell}\cdot\tilde{k}}{L}\right)^{n-1}\cdot\int^{x} e^{\frac{-y^2}{2v}}\cdot w_{r<}(y+ns_b,\delta\ell_2)\,dy}_{\text{distant mutation(s) acquired}} \\
& + \underbrace{\frac{r\cdot\tilde{\ell}\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v}\int_0^{\tilde{\ell}}\frac{dz}{\tilde{\ell}}\int_z^{\tilde{\ell}}\Pr(\delta\ell)\,d\delta\ell\int^{x} e^{\frac{-y^2}{2v}}\cdot w_{r>}(y+s_b,\delta\ell)\,dy}_{\text{colocalized mutation acquired}} \\
& + \underbrace{\frac{r\cdot\tilde{\ell}\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v}\int_0^{\tilde{\ell}}\frac{dz}{\tilde{\ell}}\int_z^{\tilde{\ell}}\Pr(\delta\ell)\,d\delta\ell\int^{x} e^{\frac{-z^2}{2v}}\,dz\int^{x_c}\frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}}\cdot w_{r<}(y+s_b,\delta\ell)\,dy}_{\text{focal mutation transferred alone and lands colocalized}} \\
& + \underbrace{\frac{r\cdot\tilde{\ell}\cdot e^{\frac{x^2}{2v}}}{2v}\int^{x} e^{\frac{-z^2}{2v}}\,dz\int^{x_c}\frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}}\cdot w_{r>}(y+s_b,\delta\ell_2)\,dy}_{\text{focal mutation transferred alone and lands distant}}\ .
\end{aligned}
\tag{67}
$$

It is instructive to consider the possible dominant balances of Eq. (67). The balance of the first two terms corresponds to the two contributions to the substitution rate in Eq. (66). Provided contributions from mutations that are not initially colocalized dominate the substitution rate, we will be able to neglect contributions from producing an additional colocalized mutation in Eq. (67). Provided,

$$
\log\left(\frac{U_b}{r\cdot\tilde{\ell}\cdot\tilde{k}}\right)\gtrsim 0\ , \tag{68}
$$

we will also be able to neglect the possibility that the new mutation becomes colocalized via a recombination event. As we will show, this will correspond to the largest possible recombination rate where recombination plays a subdominant role. This allows us to assume that this condition is true and focus on the contributions from producing or acquiring distant mutations and the focal mutation being transferred alone.

Location of the interference threshold We can obtain a condition for the interference threshold of $w_{r>}(x,\delta\ell_2)$ by substituting the shoulder solution in Eq. (60) into the non-negligible terms in Eq. (67) and evaluating the resulting integrals. Provided $x_c - ns_b \gg \sqrt{v}$, we obtain the condition,

$$
1 \approx \underbrace{\frac{U_b}{s_b}\cdot e^{\frac{x_{cr>}s_b}{v}-\frac{s_b^2}{2v}}}_{\text{distant mutation produced}} + \underbrace{\sum_{n=1}^{x_c/s_b}\frac{r\cdot\tilde{\ell}\cdot\tilde{k}}{2s_b}\cdot\left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)^n\cdot\frac{1}{n!}\cdot e^{\frac{nx_{cr>}s_b}{v}-\frac{n^2s_b^2}{2v}}}_{\text{distant mutation(s) acquired}} + \underbrace{\frac{r\cdot\tilde{\ell}}{2s_b}\cdot e^{\frac{x_{cr>}s_b}{v}-\frac{s_b^2}{2v}}}_{\text{focal mutation transferred alone}}\ .
\tag{69}
$$

We can further consider the possible dominant balances of Eq. (69). When the first term on the right hand side dominates over the second and third, it implies a mutation typically fixes by producing new mutations. This coincides with our assumption that most mutations compete asexually and take over on their original background. This allows us to assume that,

$$x_{cr>} \approx x_c \ , \tag{70}$$

while $\theta_r \sim \theta_a$, which becomes an exact equivalence in the limit $\tilde{R} \to 0$. When the second or third term dominates, a mutation will typically fix by acquiring mutations via recombination or jumping to a new background, and our asexual assumptions will have broken down. This ensures that the typical successful mutation that is not initially colocalized will be involved in at most one HGT event. This allows us to write the substitution rate peturbatively,

$$\theta_{r>} \approx \theta_a + \underbrace{\sum_{n=1}^{x_c/s_b} \frac{N\, r\cdot\tilde{\ell}\cdot\tilde{k}}{2}\cdot\left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)^n \cdot \int f(x)\cdot w_a(x+ns_b)\, dx}_{\text{distant mutation(s) acquired}} + \underbrace{\frac{N\, r\cdot\tilde{\ell}}{2}\int f(x)\cdot w_a(x+s_b)}_{\text{focal mutation transferred alone}}\, dx \ , \tag{71}$$

Our goal in this section will be to understand when these contributions from recombination start to influence the substitution rate. This will correspond to when the second and third terms dominate in Eq. (69). Since the substitution rate of transferred recombined mutations depends on $w_a(x+s_b)$, it is straightforward to evaluate their relative contribution,

$$\frac{\theta_{r>}}{\theta_a} \approx 1 + \text{contributions from acquiring mutation(s)} \ + \ \frac{r\cdot\ell}{U_b} \ . \tag{72}$$

To compare the first and second term, we must understand which term dominates the sum in Eq. (69). We can observe that the sum in Eq. (69) will achieve a maximum at n^*, which is given by,

$$n^* \approx \left\lceil \frac{x_c}{s_b} - \frac{1}{2} + \frac{v}{s_b^2}\log\left(\frac{\tilde{\pi}\cdot\tilde{\ell}}{L}\right) \right\rceil \ \approx \left\lceil \left(\frac{x_c}{s_b}-\frac{1}{2}\right)\cdot\left(1-\frac{1}{\eta}\right) \right\rceil , \tag{73}$$

where we have substituted iterative solutions for v_a and x_c from Eq. (45) on the right-hand side and defined the rescaled parameter,

$$\eta = \frac{\log(s_b/U_b)}{\log(L/\tilde{k}\cdot\tilde{\ell})} \ . \tag{74}$$

We can thus view n^* as a function of the two parameters x_c/s_b and η. The parameter x_c/s_b controls the amount of clonal interference in the population and allows us to directly assess the validity of the moderate speeds regime ($x_c/s_b \gtrsim 1$). The parameter η allows us to tune the balance between the rarity with which multiple mutations are colocalized on a random donor genome $\sim \log(\tilde{k}\cdot\tilde{\ell}/L)$ and the benefit those mutations provide to an acceptor genome $\sim \log(s_b/U_b)$.

Biologically plausible parameters The plausible range of values of η emerge from biological constraints on the genome and transfer lengths and our assumptions in the moderate speeds regime. Microbial genome lengths typically range between 10^5 and 10^7, which combined with our assumption that $\tilde{\ell} \ll L$, imply the denominator of η can range from approximately $\log(10^2)$ to $\log(10^5)$ with a

typical value between $\log(10^2)$ and $\log(10^3)$. The conditions of the moderate speeds regime in Eq. (36) require that $s_b \gg U_b$, which along with empirical estimates of the per base pair mutation rate in microbes, also set upper bound on the numerator in η. A typical mutation rate is approximately 10^{-9}, with a reasonable lower-bound at 10^{-11}. Combined with the requirement that $f_b \cdot L \gg 1$, U_b can be no lower than 10^{-9}. The maximum permissible value of the selection coefficient is set by the diffusion limit at 10^{-1}, suggesting the numerator has an upper-bound of 10^8. Collectively, these constraints suggest that η has a biological maximum ~ 4—although it will typically be much smaller—and minimum of $1/2$,

$$\frac{1}{2} \leq \eta \leq 4 \, . \tag{75}$$

In the simplest asymptotic limit, η will remain below one, in which case n^* will remain near one. We consider this scenario below.

8.1.1 Single mutation acquisition limit $\eta \lesssim 1$

We can show from Eq. (73) that the $n^* = 1$ provided

$$\frac{x_c}{s_b} - \frac{1}{2} \lesssim \frac{1}{1 - \frac{1}{\eta}} \, . \tag{76}$$

We refer to the bound in Eq. (76) as the *single mutation acquisition limit* since it determines when the contributions from acquiring single distant mutations will approach that of producing distant mutations. We will be able to consider asymptotically large population sizes ($x_c/s_b \gg 1$) in the single acquisition limit provided $\eta < 1$. For $\eta \gtrsim 1$, we can show that the upper-bound on x_c/s_b is weaker than the bound in Eq. (36) provided,

$$\log(s_b/U_b) \ll \log(L/\tilde{k} \cdot \tilde{\ell}) + 2 \, , \tag{77}$$

which will also allow us to consider asymptotically large population sizes for values of η that are fractionally larger than 1. For $\eta \gg 1$, we will be able to consider a maximum possible value of $x_c/s_b \sim 2$, which will always fail in the moderate speeds regime since it implies clonal interference is extremely weak. Consequently, for $\eta \gtrsim 1$, we will need to go beyond the single acquisition limit. We discuss this case in the next section.

Substitution rate of single transfers In the single acquisition limit, the substitution rate will scale as,

$$\frac{\theta_{r>}}{\theta_a} \approx 1 + \frac{r \cdot \tilde{\ell} \cdot \tilde{k}}{U_b} \, . \tag{78}$$

Breakdown of asexual dynamics This implies the critical recombination rate where the substitution rate will cease to be dominated by asexually evolving lineages is,

$$\log \left(\frac{U_b}{r \cdot \tilde{\ell}} \right)_{\text{critical}} \approx \log(\tilde{k}) \, . \tag{79}$$

Interpretation of the single transfer branching process When the single transfer limit is valid, the solution in Eq. (58) has a simple interpretation. The typical successful beneficial mutation will arise in the high fitness nose of the fitness distribution $x \sim x_c$ (Fig. 2C; middle). Upon surviving the stochastic fluctuations of genetic drift, this mutation will begin to grow deterministically while competing with other mutations in the nose. To take over, the focal mutation will typically survive in the nose for an additional x_c/v_a generations by producing or acquiring a total of $\sim \tilde{k}$ linked mutations (Fig. 2C; top). Each of these mutation and acquisition events will confer a benefit s_b and occur at respective rates U_b and $r \cdot \ell \cdot \tilde{k}$. This implies that the crossover where recombination becomes the dominant force will occur when the transfer and beneficial mutations rates balance $r \cdot \tilde{\ell} \cdot \tilde{k} \sim U_b$.

More rarely, with relative probability $\sim 1/\tilde{k}$, the focal mutation will not take over by staying in the nose. Instead, it will grow exponentially for $\sim x_c/v_a$ generations until it attains an $\mathcal{O}(1)$ frequency near the bulk of the fitness distribution. At a relative rate $r \cdot \tilde{\ell}$, the mutation can then jump back to the nose where it will take over by producing or acquiring linked mutations. This demonstrates how clonal competition allows mutations that would otherwise be purged in an asexual population to attain a high frequency and be rescued by recombination.

Trajectories of successful recombined single mutations These dynamics also allow us to understand the frequency trajectory of a successful recombined mutation. As we will discuss, previous work (55) has shown that the frequency spectrum of new mutations in the typical bulk is distributed as,

$$\rho(f_a) \approx \frac{\sin(\pi\alpha)}{\pi} \cdot \frac{(1-f)^{\alpha-1}}{f^{1+\alpha}} \, , \tag{80}$$

where $\alpha \equiv (1 - \tilde{k}^{-1})$. The rough transfer rate of these mutations will scale with their frequency $\sim f$ and the probability they are not at a large frequency in the present nose $\sim (1-f)$. This suggests the average transferred mutation will be at a small but modest frequency $\sim \tilde{k}^{-1}/2$. Moreover, as this transferred mutation grows to a large frequency its original background will typically decline to an overall frequency $\sim \tilde{k}^{-e}/2$. This leads to a typical trajectory where a successful recombined mutation attains a large overall frequency, declines below 1%, and then rapidly sweeps.

8.1.2 Beyond the single mutation acquisition limit

More generally, provided $\eta \gtrsim 1$, and $x_c - n^* s_b \gg \sqrt{v}$, we can show the n^* term will dominate provided,

$$\frac{n^* - 1}{1 - \frac{1}{\eta}} \lesssim \frac{x_c}{s_b} - \frac{1}{2} \lesssim \frac{n^*}{1 - \frac{1}{\eta}} \, . \tag{81}$$

This demonstrates that the value of n^* will sensitively depend on x_c/s_b, enabling us to consider asymptotically large values of x_c/s_b provided n^* is commensurately large.

Breakdown of asexual dynamics It also demonstrates that the critical recombination rate will be,

$$\log\left(\frac{U_b}{r \cdot \tilde{\ell}}\right)_{\text{critical}} \approx (n^* - 1) \cdot \left[\frac{x_c s_b}{v} - \frac{s_b^2}{2v}\right] \cdot \left(1 - \frac{1}{\eta}\right) - n^*(n^* - 1) \cdot \frac{s_b^2}{2v} + \log(\tilde{k}) \, . \tag{82}$$

For many empirically relevant values of N (and x_c/s_b), we will be able to limit our focus to the additional case where $n^* = 2$, which we respectively refer to as the *double mutation acquisition limit*. We are now in a position to evaluate $\theta_{r<}$ and assess when the contributions from producing an additional colocalized mutation will start to become important in Eq. (67).

8.2 Mutations that are initially colocalized

Provided a new mutation is initially colocalized, we can simplify the clone expansion solution in Eq. (58),

$$
\begin{aligned}
w_{r<}(x,\delta\ell_2) \approx & \underbrace{\frac{U_b \cdot e^{\frac{x^2}{2v}}}{v} \int^x e^{\frac{-y^2}{2v}} \cdot w_{r<}(y+s_b,\delta\ell_2)\, dy}_{\text{distant mutation produced}} + \underbrace{\frac{U_b \cdot e^{\frac{x^2}{2v}}}{v} \int_0^{\delta\ell_2} \Pr(\delta\ell)\, d\delta\ell \int^x e^{\frac{-y^2}{2v}} \cdot w_{r<}(y+s_b,\delta\ell)\, dy}_{\text{more colocalized mutation produced}} \\
& + \underbrace{\sum_n \frac{r\cdot\tilde{\ell}\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v} \cdot \left(\frac{\tilde{\ell}\cdot\tilde{k}}{L}\right)^{n-1} \cdot \int^x e^{\frac{-y^2}{2v}} \cdot w_{r<}(y+ns_b,\delta\ell_2)\, dy}_{\text{distant mutation(s) acquired}} \\
& + \underbrace{\frac{r\cdot\tilde{\ell}\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v} \int_0^{\delta\ell_2} \frac{dz}{\delta\ell_2} \int_z^{\delta\ell_2} \Pr(\delta\ell)\, d\delta\ell \int^x e^{\frac{-y^2}{2v}} \cdot w_{r<}(y+s_b,\delta\ell)\, dy}_{\text{more colocalized mutation acquired}} \\
& + \underbrace{\frac{r\cdot\delta\ell_2\cdot\tilde{k}\cdot e^{\frac{x^2}{2v}}}{2v} \int_0^{\delta\ell_2} \frac{dz}{\delta\ell_2} \int_z^{\delta\ell_2} \Pr(\delta\ell)\, d\delta\ell \int^x e^{\frac{-z^2}{2v}}\, dz \int^{x_c} \frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}} \cdot w_{r<}(y+s_b,\delta\ell)\, dy}_{\text{focal mutation transferred alone and lands colocalized}} \\
& + \underbrace{\frac{r\cdot\delta\ell_2\cdot e^{\frac{x^2}{2v}}}{2v} \int^x e^{\frac{-z^2}{2v}}\, dz \int^{x_c} \frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}} \cdot w_{r>}(y+s_b,\delta\ell_2)\, dy}_{\text{focal mutation transferred alone and lands distant}} \\
& + \underbrace{\frac{r\cdot(\tilde{\ell}-\delta\ell_2)\cdot e^{\frac{x^2}{2v}}}{2v} \int^x e^{\frac{-z^2}{2v}}\, dz \int^{x_c} \frac{e^{\frac{-y^2}{2v}}}{\sqrt{2\pi v}} \cdot w_{r<}(y+2s_b,\delta\ell_2)\, dy}_{\text{focal mutation transferred as pair and lands distant}} .
\end{aligned}
\tag{83}
$$

We can further simplify this branching process by considering the possible dominant balances. In the simplest scenario, the contributions from the focal mutation being transferred as a pair will remain negligible. In this case, $w_{r<}(x,\delta\ell_2)$ will become equivalent to $w_{r>}(x,\delta\ell_2)$ and $\theta_{r<}$ will be subdominant in Eq. (66). In the opposite limit, the contributions from the pair transferring will dominate and we will need to solve for $x_{cr}(\delta\ell_2)$, which will now depend on the transfer rate of the pair.

Location of the recombinant pair interference threshold To solve for the value of this interference threshold, we subsitute the shoulder solution into Eq. (83) keeping the terms corresponding to the pair

transferring and the production of linked mutations,

$$1 \approx \underbrace{\frac{U_b}{s_b} \cdot e^{\frac{x_{cr} s_b}{v} - \frac{s_b^2}{2v}}}_{\text{distant mutation produced}} + \underbrace{\frac{r \cdot (\tilde{\ell} - \delta\ell_2)}{2s_b} \cdot e^{\frac{2x_{cr} s_b}{v} - \frac{4s_b^2}{2v}}}_{\text{focal mutation transferred as pair}}, \tag{84}$$

where we have assumed $x_{cr} - 2s_b \gg \sqrt{v}$. Provided the first term dominates, we will have $x_{cr} \approx x_c$. This allows us to show that contributions from the pair transferring will start to dominate when

$$\log\left(\frac{U_b}{r \cdot (\tilde{\ell} - \delta\ell_2)}\right) \lesssim \frac{x_c s_b}{v} - \frac{3s_b^2}{2v}. \tag{85}$$

For comparison, when $\delta\ell_2 \lesssim \tilde{\ell}$, this threshold corresponds to recombination rates that are a factor of $\tilde{\ell}/L$ smaller than those where the acquisition of two mutations becomes important. When the condition in Eq. (85) becomes true, the dominant balance in Eq. (83) will come from the transfer of the focal mutation as part of a colocalized pair. This yields a condition for the interference threshold,

$$1 \approx \frac{r \cdot (\tilde{\ell} - \delta\ell_2)}{2s_b} \cdot \frac{1}{2} \cdot \left(\frac{x_{cr} - 2s_b}{x_{cr}}\right) \cdot e^{\frac{2x_{cr} \cdot s_b}{v} - \frac{4s_b^2}{2v}}, \tag{86}$$

from which we obtain the iterative solution,

$$x_{cr}(\delta\ell_2) \approx \frac{1}{2}\left[x_c + \frac{3s_b}{2} - \frac{v}{s_b}\log\left(\frac{r \cdot (\tilde{\ell} - \delta\ell_2)}{U_b}\right)\right]. \tag{87}$$

This implies that we can rewrite the condition that $x_{cr} - 2s_b \gg \sqrt{v}$ as,

$$\frac{x_c}{s_b} + \frac{v}{s_b^2}\log\left(\frac{U_b}{r \cdot (\tilde{\ell} - \delta\ell_2)}\right) \gg \frac{5}{2}, \tag{88}$$

which will always be satisfied since it coincides with the condition for clonal interference to be widespread.

Interpretation of the interference threshold The location of x_{cr} can be interpreted as a critical fitness threshold that determines a colocalized pair's fate. If a colocalized pair lands above x_{cr}, it will be transferred to so many backgrounds that it will become destined to take over. When a pair lands below x_{cr} it will have to be repeatedly transferred to take over. We can understand how the specific value of x_{cr} in Eq. (87) emerges by considering this process. Suppose a new mutation founds a colocalized pair with relative fitness x_{pair}. The pair will grow deterministically until it reaches the bulk, attaining a size $f_0 \sim \exp(x_{\text{pair}}^2/2v)$. Once in the bulk, the pair will spend $\sim \sqrt{2\pi/v}$ generations seeding new clones throughout the fitness distribution at a total rate $\tilde{R}'$. The original pair's fixation probability will depend on the total expected size of these seeded clones f_1 that land below x_{cr},

$$\langle f_1 \rangle \approx \underbrace{N \cdot \tilde{R}' \cdot f_0}_{\text{rate pair transferred}} \cdot \underbrace{\sqrt{\frac{2\pi}{v}}}_{\text{time in bulk}} \cdot \int^{x_c - 2s_b} \underbrace{\frac{e^{-\frac{y^2}{2v}}}{\sqrt{2\pi v}}}_{\text{prob acceptor}} \cdot \underbrace{\frac{e^{\frac{(y+2s_b)^2}{2v}}}{2Ny}}_{\text{peak size of acceptor}} \cdot \underbrace{2y}_{\text{prob establishes}} \, dy. \tag{89}$$

The pair will be expected to deterministically increase in frequency when $\langle f_1 \rangle \gtrsim f_0$. This yields the same condition in Eq. (87). We can understand how a successful mutation reaches these frequencies through repeated transfers with the solution to the branching process $w_{r<}(x, \delta\ell_2)$.

Interpretation of the branching process solution The solution to $w_r(x, \delta\ell_2)$ has a basic intuition when expressed in its time-dependent form,

$$w_{r<}(x, \delta\ell_2) \approx \underbrace{\frac{\tilde{R}'}{2} \cdot \int_0^\infty e^{xt - \frac{vt^2}{2}} dt}_{\text{number times pair transferred}} \cdot \underbrace{\int^{x_c} \frac{e^{-\frac{x^2}{2v}}}{\sqrt{2\pi v}} w_r(x + 2s_b, \delta\ell)\, dx}_{\text{fixation prob of acceptor}}, \tag{90}$$

where we have assumed $x \gtrsim \mathcal{O}(\sqrt{v})$. We refer to Eq. (90) as the 'clone expansion' for the transferred pair since it captures how the fixation probability of the original pair depends on the fixation probability of each of its descendants. Upon surviving genetic drift, the founding pair will begin to grow deterministically while HGT events found new clones that contain the pair. Unlike asexual clones which can only jump ahead by a factor of s_b when they produce linked mutations, the colocalized pair can jump to any fitness background in the population. The pair will have its greatest chance of being transferred when the clone is within $\mathcal{O}(\sqrt{v})$ of the mean fitness and has reached size $\sim \exp(x_{cr}^2/2v)$. With probability $\sim 1 - 2s_b/x_{cr}$, the colocalized pair will be transferred to a lineage with fitness $x_{cr} - 2s_b$ and land at $x \lesssim x_{cr}$. This process will repeat $\sim x_{cr}/2s_b$ times until a new clone is founded above x_{cr}. This clone will become so large that it will be guaranteed to take over.

Trajectories of successful colocalized pairs These dynamics and the successful trajectory of a colocalized pair have several interesting features. First, the focal mutation that founds the colocalized pair (the second mutation in the pair) will arise on a random genetic background drawn from the fitness class at $x_c - s_b$. Typically, this background and the first mutation in the pair will be at detectable frequency ($\gg 1\%$) within this fitness class. In contrast, the founding mutation in the pair will land at fitness $x_{cr} \lesssim x_c$, where it will compete against a fitness class that has already been growing exponentially. As a result, this second mutation will represent an extremely small fraction ($\ll 1\%$) of the fitness class at x_{cr}. This means that the first mutation in the pair will reach a large overall frequency when its fitness class reaches the bulk, while the second mutation will remain undetectable by metagenomic sequencing standards. Since the two mutations take over as a pair, the first mutation will then decline in frequency over $\sim x_c/s_b$ mutations until it too is undetectable.

This lands us at the second interesting feature: the colocalized pair will remain at undetectable frequency for at least as long as the sweep time. Each of the $x_{cr}/2s_b$ jumps back to x_{cr}, the pair must wait $\sim x_{cr}/v$ generations before it reaches the bulk and can transfer again. This process will last $\sim x_{cr}/2s_b \cdot x_{cr}/v$ generations until the pair can transfer to the nose at $\sim x_c$ and take over by producing linked mutations. As a result a colocalized pair will remain at small frequency for a period that depends on the underlying recombination rate,

$$\frac{x_c}{v} \lesssim \frac{1}{s_b} \cdot \frac{x_{cr}^2}{2v} \lesssim \frac{1}{2} \cdot \frac{x_c}{s_b} \cdot \frac{x_c}{v}, \tag{91}$$

where x_c/v is approximately the fixation time of an asexual mutation. This implies that the typical fixation time of a colocalized pair is at least twice as large as that of the typical successful mutation.

Extending the shoulder solution The clone expansion solution will remain valid provided $x > x_{\text{min,r}}$, where

$$x_{\text{min,r}} \equiv \mathcal{O}(\sqrt{v})\ . \tag{92}$$

This value is far smaller than $x_{\min,\mathrm{a}} = x_c - s_b$, which suggests that the fixation probability of colocalized pairs $w_{r<}(x, \delta\ell_2)$ will decay far slower that that of mutations that fix by producing linked mutations $w_a(x)$. This highlights that mutations can have a far larger chance of taking over when they arise on poor genetic backgrounds—an important advantage of recombination. When the pair arises on a value of x below $\mathcal{O}(\sqrt{v})$, we obtain a different solution to $w_{r<}(x, \delta\ell_2)$,

$$w_{r<}(x, \delta\ell_2) = \sqrt{\frac{2v}{\pi}} \cdot \frac{x_{cr}}{|x|} e^{-\frac{x_{cr}^2}{2v}} \ . \tag{93}$$

This solution also has a simple interpretation. A colocalized pair arising with a negative fitness compared to the mean will fluctuate at small frequency. To take over, it will need to jump to a background with positive fitness, where it can then grow exponentially.

Substitution rate We are now in a position to evaluate the contributions to the substitution rate from colocalized pairs,

$$\theta_{r<} = NU_b \int_0^{\tilde{\ell}} \frac{d\delta\ell_2}{\tilde{\ell}} \int f(x) \cdot w_{r>}(x + s_b, \delta\ell_2)\, dx \ , \tag{94}$$

which are founded by the second mutation that carries a fitness benefit of s_b. We can substitute the solution for $w_{r<}(x, \delta\ell_2)$ and $w_a(x)$ to show that,

$$\frac{\theta_{r<}}{\theta_a} - 1 \approx \int_0^{\tilde{\ell}} \tilde{k} \cdot \frac{d\delta\ell_2}{L} \cdot \frac{x_{cr}^2}{x_c s_b} \cdot e^{\frac{(x_c - s_b)^2}{2v} - \frac{(x_{cr} - s_b)^2}{2v}} \ . \tag{95}$$

The value of this integral will depend on $\delta\ell_2$ through $x_{cr2}(\delta\ell_2)$. Substituting the value of x_{cr} in Eq. (87), we obtain,

$$\frac{\theta_{r<}}{\theta_a} - 1 \approx e^{\frac{3}{4}\cdot\frac{x_c^2}{2v} - \frac{7}{8}\cdot\frac{x_c s_b}{v} + \frac{15 s_b^2}{32v}} \cdot \int_0^{\tilde{\ell}} \tilde{k} \cdot \frac{d\delta\ell_2}{L} \cdot e^{\left(\frac{1}{4}\cdot\frac{x_c}{s_b} - \frac{1}{8}\right)\cdot\log\left(\frac{r\cdot(\tilde{\ell} - \delta\ell_2)}{U_b}\right) - \frac{1}{8}\cdot\frac{v}{s_b^2}\log^2\left(\frac{r\cdot(\tilde{\ell} - \delta\ell_2)}{U_b}\right)} \ . \tag{96}$$

The integrand in Eq. (96) will remain constant provided $\delta\ell_2 \lesssim \tilde{\ell}/2$ before declining to zero. This allows us to approximate the integral up to an $\mathcal{O}(1)$ constant,

$$\frac{\theta_{r<}}{\theta_a} - 1 \approx \frac{\tilde{k}\cdot\tilde{\ell}}{L} \cdot e^{\frac{3}{4}\cdot\frac{x_c^2}{2v} - \frac{7}{8}\cdot\frac{x_c s_b}{v} - \left(\frac{1}{4}\cdot\frac{x_c}{s_b} - \frac{1}{8}\right)\cdot\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right) + \frac{15 s_b^2}{32v} - \frac{1}{8}\cdot\frac{v}{s_b^2}\log^2\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)} \ . \tag{97}$$

We can subsitute the iterative solutions for v_a and x_c to obtain the first order approximation,

$$\frac{\theta_{r<}}{\theta_a} - 1 \approx \left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right) \cdot (NU_b)^{\frac{3}{4}} \cdot \left(\frac{r\cdot\tilde{\ell}}{U_b}\right)^{\frac{1}{4}\cdot\frac{x_c}{s_b} - \frac{1}{8}} \cdot \left(\frac{U_b}{s_b}\right)^{\frac{1}{8}} \ . \tag{98}$$

Breakdown of asexual dynamics To understand when the substitution rate of colocalized pairs come to dominate the overall fixation probability, we set Eq. (97) equal to one. This yields,

$$\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)_{\text{critical}} \approx \frac{2x_c s_b}{v} \cdot \left[1 - \frac{2s_b}{x_c} + \frac{s_b^2}{x_c^2} + \frac{2v}{x_c^2}\log\left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)\right]^{1/2} - \frac{x_c s_b}{v} + \frac{s_b^2}{2v} \ , \tag{99}$$

which can be expressed in terms of η,

$$\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)_{\text{critical}} \approx \frac{2x_c s_b}{v}\cdot\left[1-\left(1+\frac{1}{\eta}\right)\left(\frac{2s_b}{x_c}-\frac{s_b^2}{x_c^2}\right)\right]^{1/2} - \frac{x_c s_b}{v}+\frac{s_b^2}{2v}\ . \tag{100}$$

We can expand around $x_c/s_b \gg 1$, to find that the leading order behavior of Eq. (100) will be

$$\log\left(\frac{U_b}{r\cdot\ell}\right)_{\text{critical}} \sim \left(1-\frac{s_b}{x_c}\right)\cdot\log\left(\frac{s_b}{U_b}\right) - 2\cdot\frac{s_b}{x_c}\cdot\log\left(\frac{L}{\tilde{k}\cdot\tilde{\ell}}\right)\ , \tag{101}$$

which approaches

$$\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)_{\text{critical}} \sim \log\left(\frac{s_b}{U_b}\right)\ , \tag{102}$$

for large values of x_c/s_b. We can also substitute Eq. (101) into x_{cr} to show that the smallest possible value of x_{cr} will be,

$$x_{cr} \geq x_c - s_b\cdot\left(1-\frac{1}{\eta}\right)\ . \tag{103}$$

This ensures that colocalized pairs will always arise above $x_c - s_b$ since we have assumed $\eta \gtrsim 1$.

8.3 Assessing the dominant contribution from recombination

In the previous sections we have outlined several processes from recombination that contribute to the fixation of a new mutation. We are now in a position to compare the critical recombination rates associated with these processes and ask which one actually causes asexuality to breakdown.

Comparison to single acquisition limit Provided $n^* = 1$ and $x_c/s_b \lesssim 1/(1-1/\eta)+1/2$, we can compare the critical rates in Eq. (100) and Eq. (79). For these values, we find the asexual assumption will break down due to the repeated transfer of colocalized pairs as opposed to acquisitions of single mutations provided,

$$\frac{x_c}{s_b} \gtrsim \frac{1}{3\cdot\eta}\cdot\left[\left(\sqrt{\eta^2+16\cdot\eta+16}+4+\frac{7}{2}\cdot\eta\right)\right]\ . \tag{104}$$

We can also show this will occur only while $x_{cr}-2s_b \gg \sqrt{v}$ by substituting Eq. (100) into Eq. (88) to obtain a condition on x_c and η that is weaker than the one in Eq. (104). We can also show that this condition will be valid up to values of $\eta = 3/2$ by setting Eq. (104) equal to $1/(1-1/\eta)+\frac{1}{2}$. This allows us to expand Eq. (104) to first order around $\eta = 0$, to obtain the approximation,

$$\frac{x_c}{s_b} \gtrsim \frac{8}{3\cdot\eta}+\frac{11}{6} \iff \frac{3}{4}\cdot\log\left(NU_b\right) \gtrsim \log\left(\frac{L}{\tilde{\ell}\cdot\tilde{k}}\right) \gtrsim \frac{2}{3}\cdot\log\left(\frac{s_b}{U_b}\right)\ . \tag{105}$$

where the right hand side is obtained by substituting iterative approximations for x_c. For larger values of η we will need to compare Eq. (100) to the double acquisition limit.

Comparison to double acquisition limit Provided $1/(1-1/\eta)+1/2 \lesssim x_c/s_b \lesssim 2/(1-1/\eta)+1/2$, n^* will be equal to 2 and we can compare Eq. (100) and Eq. (82). This comparison shows that the asexual assumption will break down due to the repeated transfer of colocalized pairs provided

$$\frac{x_c}{s_b} \gtrsim \frac{24\eta - 1}{8\eta - 2} \, , \tag{106}$$

which can be approximated by a first order expansion around $\eta \gtrsim 1$,

$$\frac{x_c}{s_b} \gtrsim 3 + \frac{5}{8 \cdot \eta} \iff \log(NU_b) - \frac{1}{4} \cdot \log\left(\frac{L}{\tilde{\ell} \cdot \tilde{k}}\right) \gtrsim \log\left(\frac{s_b}{U_b}\right) \gtrsim \frac{3}{2} \log\left(\frac{L}{\tilde{\ell} \cdot \tilde{k}}\right) \, . \tag{107}$$

We can show Eq. (106) will remain valid up to $\eta = 4$ by setting it equal to the condition $2/(1-1/\eta)+1/2$ and solving for η. This ensures that we will be able to ignore the contributions from $n^* > 2$. It also suggests that the double acquisition limit will only apply for a narrow boundary $2 + \sqrt{v}/s_b \lesssim x_c/s_b \lesssim 3 + 5/8 \cdot \eta$. In practice, we will be able to neglect this boundary region since since $s_b \sim \sqrt{v}$ for many empirical scenarios.

It is helpful to briefly summarize the results thus far. Due to largely asexual competition and clonal interference, any genotype will carry multiple ($\sim x_c/s_b$) unique beneficial mutations. A new mutation will arise on one of these genotypes and typically land distant from any other preexisting mutations. It will then compete against the other individuals in the population largely by producing linked mutations. More rarely, the focal mutation can acquire a single mutation via a transfer event. This transferred mutation will typically land distant from the focal mutation. This process acts like an effective mutation rate, which allows the mutation to move ahead by a factor of s_b as it would by producing linked mutations. Alternatively, the new mutation could also acquire multiple genetically distant colocalized mutations. Simply by chance, some of the mutations on a donor genotype will be colocalized in pairs (or tuplets) allowing them to be transferred together. These colocalized mutations can attain appreciable frequencies simply by chance in asexual populations by occurring on high fitness background. The new mutation does not care which colocalized pair it acquires, so the overall rate of acquiring distant colocalized pairs will depend on the total number of colocalized pairs in the population and their respective frequency. This process also acts like an effective mutation rate with a $2s_b$ benefit and implies that a colocalized pair takes over after a single transfer.

From the perspective of any specific colocalized pair, it can also take over by *repeatedly* being transferred to more fit genetic backgrounds and growing in frequency. Rarely, a new mutation will arise near another preexisting mutation and become part of one of these pairs. As we have shown, this pair can be most likely to take over by repeatedly switching backgrounds well before quasi-asexuality breaks down. We emphasize that this picture is consistent with the above breakdown in quasi-asexuality due to multiple separate acquisition events. Since $x_{cr} < x_c$, any *specific* colocalized pair can still be extremely unlikely to be transferred above x_{cr} —where it would not need to be transferred again to take over— even while the total chance that *any* colocalized pair being transferred above x_{cr} is much larger.

Crucially, these solutions required us to to be able to ignore contributions from the productions of colocalized triplets. We are now in a position to evaluate the substitution rate of these triplets and determine when this approximation will be valid.

8.4 Mutations arising as colocalized triplets

While we have thus far focused on the possibility that only two mutations are transferred together, three mutations can also become colocalized. Provided $\theta_r/\theta_a - 1 \lesssim 1$, we can assume $r \cdot \tilde{\ell} \lesssim U_b$ so that these triplets will continue to be founded by mutation. Since we have shown colocalized pairs will remain at small frequency until they are destined to take over, we can assume the abundance of colocalized pairs is proportional to $(\tilde{\ell}/L)^2$. This allows us to extend θ_r,

$$\theta_r \approx \underbrace{\theta_{r>}}_{\text{not colocalized}} + \underbrace{\theta_{r<2}}_{\text{colocalized pair}} + \underbrace{NU_b \int \Pr(\delta\ell_3)\, d\delta\ell_3 \cdot \int f(x) \cdot w_{r>}(x + s_b, \delta\ell_{23})\, dx}_{\theta_{r<3}\text{ colocalized triplet}}, \tag{108}$$

where $\Pr(\delta\ell_3)$ is defined in Eq. (55).

Clone expansion solution for colocalized triplets Provided the contributions from colocalized triplets become relevant before asexuality has broken down, we can express $w_r(x, \delta\ell_3)$ as,

$$\begin{aligned} w_{r<}(x, \delta\ell_3) &= \underbrace{\frac{U_b \cdot e^{\frac{x^2}{2v}}}{v} \cdot \int_{-\infty}^{x} e^{-\frac{y^2}{2v}} \cdot w_{r<}(y + s_b, \delta\ell_3)\, dy}_{\text{distant mutation produced}} \\ &+ \underbrace{\frac{r \cdot e^{\frac{x^2}{2v}}}{2v} \cdot \int_{-\infty}^{x} e^{-\frac{z^2}{2v_a}} dz \cdot \int_{-\infty}^{x_c} \frac{e^{-\frac{y^2}{2v_a}}}{\sqrt{2\pi v_a}}\, dy \cdot \int_{\delta\ell_3}^{\tilde{\ell}} d\ell\ \lambda(\ell) \cdot (\ell - \delta\ell_3) \cdot w_{r,3}(y + 3s_b, \delta\ell_3)}_{\text{triplet transferred and lands distant from acceptor mutations}}, \end{aligned} \tag{109}$$

This solution will depend on whether the triplet will need multiple transfers or only a single transfer to take over. The fixation probability must eventually saturate to a linear scaling when the number of colocalized mutations is so large that a single transfer ensures their fixation. This will occur for colocalized triplets provided $3s_b - x_{cr<} \gtrsim \sqrt{v}$, where $x_{cr<}$ depends on the underlying transfer rate. When $x_c/s_b \lesssim 3$, this will always be true. For larger values of x_c/s_b, we will have $x_{cr<} - 3s_b \gg \sqrt{v}$ and this crossover will occur at a characteristic transfer rate. We study these cases separately below.

8.4.1 Single transfer limit

Provided $3s_b - x_{cr<} \gg \sqrt{v}$, we can substitute the shoulder solution into Eq. (109) to obtain a condition for the interference threshold $x_{cr<}$,

$$1 \approx \underbrace{\frac{U_b}{s_b} \cdot \frac{x_{cr<} - s_b}{x_{cr<}} \cdot e^{\frac{x_{cr<} s_b}{v} - \frac{s_b^2}{2v}}}_{\text{production of distant mutations}} + \underbrace{\frac{r \cdot (\tilde{\ell} - \delta\ell_3) \cdot s_b \cdot \sqrt{2\pi v} e^{\frac{x_{cr<}^2}{2v}}}{v x_{cr<}}}_{\text{triplet transferred}}, \tag{110}$$

Provided,

$$\log\left(\frac{U_b}{r \cdot (\tilde{\ell} - \delta\ell_3)}\right) \gtrsim \frac{(x_{cr<} - s_b)^2}{2v}, \tag{111}$$

contributions from the triplet transferring will dominate over those from producing linked mutations. In this case the interference threshold will be solely determined by contributions from the triplet transferring and we can obtain an iterative approximation for $x_{cr<}$,

$$x_{cr<} \approx \sqrt{2v} \cdot \log^{1/2}\left(\frac{s_b}{r\cdot(\tilde{\ell}-\delta\ell_3)}\right) = \sqrt{(2x_c - s_b)s_b - 2v\cdot\log\left(\frac{r\cdot(\tilde{\ell}-\delta\ell_3)}{U_b}\right)}\,. \tag{112}$$

This approximation will be valid while $3s_b - x_{cr<} \gtrsim \sqrt{v}$, which implies

$$5 \gtrsim \frac{x_c}{s_b} + \frac{v}{s_b^2}\cdot\log\left(\frac{U_b}{r\cdot(\tilde{\ell}-\delta\ell_3)}\right)\,. \tag{113}$$

Substitution rate of colocalized triplets We are now in a position to evaluate the substitution rate of colocalized triplets. We can substitute our solutions for $w_{r,3}(x,\delta\ell_3)$ and $w_a(x)$ to show that,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \approx \left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)^2 \cdot \int_0^{\tilde{\ell}} \frac{6\cdot d\delta\ell_3}{\tilde{\ell}}\cdot\left(\frac{\delta\ell_3}{\tilde{\ell}}\right)\cdot\left(1-\frac{\delta\ell_3}{\tilde{\ell}}\right)\cdot\frac{x_{cr<}^2}{x_c s_b}\cdot e^{\frac{(x_c-s_b)^2}{2v} - \frac{(x_{cr<}-s_b)^2}{2v}}. \tag{114}$$

Substituting solutions for $x_{cr<3}$ and x_c, we obtain,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \approx \left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)^2 \cdot e^{\frac{x_c^2}{2v} - \frac{2x_c s_b}{v} - \frac{s_b^2}{2v} + \log\left(\frac{r\cdot\tilde{\ell}}{U_b}\right)} \int_0^{\tilde{\ell}} \frac{6\cdot d\delta\ell_3}{\tilde{\ell}}\cdot\left(\frac{\delta\ell_3}{\tilde{\ell}}\right)\cdot\left(1-\frac{\delta\ell_3}{\tilde{\ell}}\right)^2 \cdot e^{\sqrt{\frac{2s_b^2}{v}\left(\log\left(\frac{s_b}{r\cdot\tilde{\ell}}\right) - \log\left(1-\frac{\delta\ell_3}{\tilde{\ell}}\right)\right)}}\,. \tag{115}$$

The integrand will remain roughly constant from $\delta\ell_3 \lesssim \tilde{\ell}/2$, allowing us to substitute values for x_c and v to approximate,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \approx \left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right)^2 \cdot \left(N\cdot r\cdot\tilde{\ell}\right)\cdot\left(\frac{U_b}{s_b}\right)^{\left[1-\sqrt{\frac{\log(s_b/r\cdot\tilde{\ell})}{\log(Ns_b)}}\right]}\cdot e^{\frac{s_b^2}{2v}}\,. \tag{116}$$

Breakdown of asexuality It is now straightforward to ask when single transfers of colocalized triplets will cause asexuality to break down. The critical recombination rate can be obtained by setting the left-hand side of Eq. (114) equal to one,

$$\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)_{\text{critical}} \sim \frac{x_c s_b}{v}\cdot\left[1-\left(1+\frac{2}{\eta}\right)\cdot\left(\frac{2s_b}{x_c} - \frac{s_b^2}{x_c^2}\right)\right]^{1/2} - 2\cdot\left(1+\frac{1}{\eta}\right)\left[\frac{x_c s_b}{v} - \frac{s_b^2}{2v}\right] + \frac{s_b^2}{2v} + \frac{x_c^2}{2v}\,. \tag{117}$$

We can now ask when this critical recombination rate becomes smaller than that from the repeated transfer of colocalized pairs in Eq. (100). For $1/2 \le \eta \le 4$, we can show that this crossover will always occur when the assumption that $3s_b - x_{cr<} \gg \sqrt{v}$ in Eq. (113) is no longer valid. This means colocalized triplets will always take over via multiple transfers, which implies $x_{cr<} - 3s_b \gg \sqrt{v}$. We consider this case below.

8.4.2 Repeated transfers of colocalized triplets

Provided $x_{cr<} - 3s_b \gg \sqrt{v}$, the location of the interference threshold takes the form,

$$1 \approx \left(\frac{U_b}{s_b}\right) \cdot \frac{x_{cr<} - s_b}{x_{cr<}} \cdot e^{\frac{x_{cr<} \cdot s_b}{v} - \frac{s_b^2}{2v}} + \left(\frac{r \cdot (\tilde{\ell} - \delta\ell_3)}{2s_b}\right) \cdot \frac{1}{3!} \cdot \frac{x_{cr<} - 3s_b}{x_{cr<}} \cdot e^{\frac{3x_{cr<} \cdot s_b}{v} - \frac{9s_b^2}{2v}} \tag{118}$$

.

Provided,

$$\log\left(\frac{U_b}{r \cdot (\tilde{\ell} - \delta\ell_3)}\right) \lesssim \frac{2x_c s_b}{v} - \frac{4s_b^2}{2v}\,, \tag{119}$$

the contributions from the triplet transferring will dominate over those from mutation. This allows us to approximate the location of the interference threshold as, with $x_{cr<}$ given by,

$$x_{cr<} \approx \frac{1}{3}\left[x_c + 4s_b - \frac{v}{s_b}\log\left(\frac{r \cdot (\tilde{\ell} - \delta\ell_3)}{U_b}\right)\right]\,. \tag{120}$$

Substitution rate of colocalized triplets We can substitute this value of $x_{cr<}$ into Eq. (114) to calculate the substitution rate of colocalized triplets,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \approx \left(\frac{\tilde{k} \cdot \tilde{\ell}}{L}\right)^2 \cdot e^{\frac{4}{9}\cdot\frac{x_c^2}{2v} - \frac{10}{9}\cdot\frac{x_c s_b}{v} + \frac{4}{9}\cdot\frac{s_b^2}{v}} \cdot \int_0^{\tilde{\ell}} \frac{6 \cdot d\delta\ell_3}{\tilde{\ell}} \cdot \left(\frac{\delta\ell_3}{\tilde{\ell}}\right) \cdot \left(1 - \frac{\delta\ell_3}{\tilde{\ell}}\right) \cdot e^{-\frac{1}{9}\left[\frac{x_c}{s_b}+1\right]\cdot\log\left(\frac{U_b}{r\cdot(\tilde{\ell}-\delta\ell_3)}\right) - \frac{v}{18s_b^2}\cdot\log^2\left(\frac{U_b}{r\cdot(\tilde{\ell}-\delta\ell_3)}\right)}\,. \tag{121}$$

Provided $\delta\ell_3 \lesssim \tilde{\ell}/2$, the integrand will remain roughly constant. This allows us to approximate the integral,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \approx \left(\frac{\tilde{k} \cdot \tilde{\ell}}{L}\right)^2 \cdot e^{\frac{4}{9}\cdot\frac{x_c^2}{v} - \frac{10}{9}\cdot\frac{x_c s_b}{v} + \frac{4}{9}\cdot\frac{s_b^2}{v} - \frac{1}{9}\left[\frac{x_c}{s_b}+1\right]\cdot\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right) - \frac{v_a}{18s_b^2}\cdot\log^2\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)}\,. \tag{122}$$

We can substitute iterative solutions for x_c and v_a to obtain,

$$\frac{\theta_{r<3}}{\theta_a} - 1 \sim \left(\frac{\tilde{k} \cdot \tilde{\ell}}{L}\right)^2 \cdot (NU_b)^{\frac{8}{9}} \cdot \left(\frac{U_b}{s_b}\right)^{\frac{2}{9}} \cdot \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)^{\frac{1}{9}\cdot\frac{x_c}{s_b}}\,, \tag{123}$$

where we have simplified to first order in $x_c - s_b \gg \sqrt{v}$.

8.4.3 Breakdown of asexuality

We are now in a position to ask when quasi-asexuality will break down due to repeated transfers of colocalized triplets. To do this, we first calculate the critical recombination rate by setting Eq. (122) equal to one to obtain,

$$\log\left(\frac{U_b}{r \cdot \tilde{\ell}}\right)_{\text{critical}} \sim \frac{3x_c s_b}{v} \cdot \left[1 - \left(1 + \frac{2}{\eta}\right) \cdot \left(\frac{2s_b}{x_c} - \frac{s_b^2}{x_c^2}\right)\right]^{1/2} - \frac{x_c s_b}{v} - \frac{s_b^2}{v}\,. \tag{124}$$

We can now compare this critical rate to those in Eq. (104) and Eq. (106). First for $1/2 \leq \eta \leq 3/2$, we can compare Eq. (124) and Eq. (104), to show that colocalized triplets will begin to dominate when,

$$\frac{3x_c}{s_b}\left[1-\left(1+\frac{2}{\eta}\right)\cdot\left(\frac{2s_b}{x_c}-\frac{s_b^2}{x_c^2}\right)\right]^{1/2}-\frac{3}{2}-\frac{2x_c}{s_b}\left[1-\left(1+\frac{1}{\eta}\right)\cdot\left(\frac{2s_b}{x_c}-\frac{s_b^2}{x_c^2}\right)\right]^{1/2}\gtrsim 0\,. \tag{125}$$

To first order in $x_c/s_b \gg 1$ and $\eta \ll 1$, this yields the approximate condition,

$$\frac{x_c}{s_b} \gtrsim \frac{5}{\eta}+\frac{29}{12}\,. \tag{126}$$

For $3/2 \leq \eta \leq 4$, we can set Eq. (124) equal to Eq. (106) to show that

$$\frac{x_c}{s_b} \gtrsim \frac{1}{64}\left(104-\frac{20}{1-4\eta}+\frac{144}{\eta}+\frac{12\sqrt{144-1048\eta+1337\eta^2+1904\eta^3+1600\eta^4}}{\sqrt{\eta^2-8\eta^3+16\eta^4}}\right)\,, \tag{127}$$

which can be expanded around $\eta \gg 1$ to yield the approximate condition

$$\frac{x_c}{s_b} \gtrsim \frac{7}{2}+\frac{313}{80\cdot\eta}\,. \tag{128}$$

Substituting the iterative solution for x_c, we can show this implies,

$$NU_b \gtrsim \left(\frac{s_b}{U_b}\right)\cdot\left(\frac{L}{\tilde{k}\cdot\tilde{\ell}}\right)^2\,. \tag{129}$$

For these large population sizes, we can show that quasi-asexuality will break down when,

$$\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)_{\text{critical}} \sim \left(\frac{s_b}{U_b}\right)^{2\cdot(1-1/\tilde{k})}\cdot\left(\frac{\tilde{\ell}\cdot\tilde{k}}{L}\right)^{2/\tilde{k}}\,. \tag{130}$$

8.5 Larger tuplet sizes

In larger populations, the contributions from larger tuplet sizes like colocalized quadruplets will become important. Provided these contributions continue to be dominated by multiple transfers, we can straightforwardly extend our approach to a tuplet size j. This will yield a condition for the interference threshold,

$$x_{cr<} \approx \frac{1}{j}\left[x_c+\frac{s_b}{2}\left(j^2-1\right)+\frac{v}{s_b}\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)\right]\,, \tag{131}$$

which can be used to evaluate the substitution rate.

8.6 Rate of adaptation

While we have focused on quasi-asexual dynamics where the rate of adaptation is roughly equal to the asexual value, we can straightforwardly calculate the corrections to this limit. As recombined mutations begin to reach a large frequency the width of the fitness distribution and rate of adaptation will begin to shift from their asexual values,

$$v \approx v_a+\delta v\,, \tag{132}$$

$$x_c \approx x_{ca} + \delta x_c. \tag{133}$$

To estimate v we can extend the self-consistency calculation for the rate of adaptation

$$v \approx \underbrace{NU_b \cdot \int f(x) \cdot s_b \cdot w_a(x+s_b)\ dx}_{\text{production of linked mutations}} + \underbrace{Nr \cdot \tilde{\ell} \cdot \tilde{k} \int f(x) \cdot s_b \cdot w_a(x+s_b)\ dx}_{\text{transfer of single mutations}} \tag{134}$$

$$+ \underbrace{NU_b \cdot \frac{\tilde{\ell} \cdot \tilde{k}}{L} \int f(x) \cdot s_b \cdot w_{r2}(x+s_b)\ dx}_{\text{production of colocalized pairs}} + \underbrace{NU_b \cdot \left(\frac{\tilde{\ell} \cdot \tilde{k}}{L}\right)^2 \int f(x) \cdot s_b \cdot w_{r3}(x+s_b)\ dx}_{\text{production of colocalized triplets}} . \tag{135}$$

We can evaluate these integrals to approximate,

$$1 \approx \frac{2NU_b}{\sqrt{2\pi v}} \left[\left(1 + \frac{r \cdot \tilde{\ell} \cdot \tilde{k}}{U_b}\right) \cdot x_{ca} e^{-\frac{(x_c - s_b)^2}{2v}} + \left(\frac{\tilde{\ell} \cdot \tilde{k}}{L}\right) \cdot x_{cr2} e^{-\frac{(x_{cr2} - s_b)^2}{2v}} + \left(\frac{\tilde{\ell} \cdot \tilde{k}}{L}\right)^2 \cdot x_{cr3} e^{-\frac{(x_{cr3} - s_b)^2}{2v}} \right] . \tag{136}$$

In small populations where the first order corrections originate from single transfers the new conditions for x_c and v are identical to the asexual ones for x_{ca} and v_a with the effective parameter $U_b \to U_b + r \cdot \tilde{\ell} \cdot \tilde{k}$. This allows us to show the rate of adaptation will initially increase as,

$$v \approx 2s_b^2 \cdot \frac{\log(N\sqrt{s_b \cdot (U_b + r \cdot \tilde{\ell} \cdot \tilde{k})})}{\log(s_b/(U_b + r \cdot \tilde{\ell} \cdot \tilde{k})} . \tag{137}$$

For larger pairs, when colocalized pairs start to dominate the overall fixation rate, we can show when quasi-asexuality breaks down,

$$v/v_a - 1 \sim 1/\log(Ns_b) . \tag{138}$$

This demonstrates that recombined mutations can make up an appreciable number of overall fixations even while the rate of adaptation shifts modestly from its asexual value.

9 Asexual genetic diversity and evolutionary trajectories

When the typical mutation takes over via repeated transfer events, multiple transferred mutations will reach high frequency concurrently. As a result, the quasi-asexual regime will breakdown since we can no longer assume that a transferred mutation competes against a predominantly asexual population. In the simplest case, we would be able to extend our continuous traveling wave framework and self-consistently calculate the rate of fixation and adaptation based on this competition among recombined mutations. However, we can already see that this approach will be more complicated than the steady state picture we have considered thus far. In particular, the rate at which new recombined clones are founded will depend on the number of mutations unique to a randomly sampled donor. This value can vary over time causing non-equilibrium fluctuations in the fitness distribution and rate of adaptation. As a result, the linear approximation to the change in the mean fitness in Eq. (22) will fail and the shape of the fitness distribution will no longer maintain a steady shape.

To understand these fluctuations, we must understand the feedback between the high-fitness nose where new mutations that reach large frequency first arise and the bulk where most of the donated

mutations originate. To do this, we return to the picture of discrete fitness classes. This allows us to keep track of genetic diversity by analyzing how extremely fit individuals arise in the nose and trace the dynamics of adaptation forward in time. Our approach will also rely heavily on previous work modeling asexual genetic diversity (56)(55)(35)(57)(33). We briefly review key results before showing how they can be extended in sexual populations.

9.1 Stochastic dynamics in the nose

In the fitness class picture, the distribution of fitness in the population can be coarse-grained into discrete classes with relative fitness $x = n \cdot s_b$,

$$\sum_n f(n \cdot s_b) = 1 \tag{139}$$

This picture is useful because it allows us to consider the genetic diversity in each class separately. Each fitness class is composed of clones or mutations with frequency g_i,

$$\sum_{i=1}^{L \cdot f_b} h(g_i|x) = 1 \ . \tag{140}$$

In an asexual population, exactly one mutation in each fitness class will be linked to the eventual common ancestor and take over. Since mutations in the same fitness class have the same fitness, they are effectively neutral with respect to one another. This implies that a mutation's chance of taking over is simply proportional to its frequency within the fitness class it was founded. Successful mutations typically arise in the high-fitness nose of the population where they have the greatest chance of reaching large frequencies. These mutations arise with a relative fitness $x = x_{ca}(t)$ and are produced by a feeding class directly below the nose $x = x_{ca}(t) - s_b$. Since $x_{ca}(t) - s_b \gg \sqrt{v_a}$, the frequency of the nose will remain small $f(x_{ca}, t) \ll 1$ such that it can be described with a linear branching process approximation of Eq. (20),

$$\frac{\partial f(x_{ca}, t)}{\partial t} = \underbrace{x_{ca}(t) \cdot f(x_{ca})}_{\text{selection}} + \underbrace{U_b \cdot f(x_{ca} - s_b, t)}_{\text{mutation}} + \underbrace{\sqrt{\frac{f(x_{ca})}{N}} \eta(t)}_{\text{genetic drift}} \ , \tag{141}$$

where $\eta(t)$ is a standard Brownian noise term. While the exact fitness of the nose will depend on the advance of the mean fitness, previous work has shown a reasonable approximation is to assume that the fitness of the nose remains constant while it establishes $x_{ca}(t) \approx x_{ca}$ (33). This constant value corresponds to the steady state value we solve for in the traveling wave model. We define $t_0 = 0$ as the time when the class at $f(x_{ca} - s_b)$ first establishes and $f(x_{ca} - s_b, 0) = 1/N x_{ca}$, so it can be described as exponentially growing,

$$\underbrace{f(x_{ca} - s_b, t)}_{\text{size of feeding class}} = \frac{e^{\int_{t_0}^{t} dt' (x_{ca}(t') - s_b)}}{N x_{ca}} \ . \tag{142}$$

The nose class will establish τ_q generations after t_0 and begin to grow exponentially.

$$\underbrace{f(x_{ca}, t)}_{\text{size of nose}} = \nu(t) \cdot \frac{e^{\int_{\tau_q}^{t} dt' x_{ca}(t')}}{N x_{ca}} \ , \tag{143}$$

where $\nu(t)$ is a random variable determined by the stochastic dynamics in Eq. (141) that converges at long times.

9.2 Fate of individual variants: linked selection

The fate of any given mutation g_j in the feeding class will immediately depend on the linked descendants it produces in the next class. The number of descendants will in turn depend on the mutation's overall frequency in the feeding class $h(g_j|x_{ca} - s_b)$,

$$\frac{\partial h(g_j|x_{ca},t)}{\partial t} = x_{ca}(t) \cdot h(g_j|x_{ca}) + U_b \cdot h(g_j|x_{ca} - s_b, t) \cdot f(x_{ca} - s_b, t) + \sqrt{\frac{h(g_j|x_{ca})}{N}}\eta(t) \ . \quad (144)$$

Previous work (55) has shown that we can track the fate of this mutation over subsequent establishments using the dynamics in Eq. (144). The probability of finding the mutation at $h(g_j|x_{ca}) \equiv h$ in the next nose given its frequency in the feeding class $h(g_j|x_{ca} - s_b) \equiv h_0$ is described by the kernel $\rho_a(h|h_0)$. This kernel arises from the stochastic competition between two classes—one with the mutation and one without it— each feeding linked beneficial mutants into the new nose at rates θ_0 and θ_1,

$$\underbrace{\theta_0 = U_b \cdot h_0}_{\text{feeding rate with mutation}} \quad , \quad (145)$$

$$\underbrace{\theta_1 = U_b \cdot (1 - h_0)}_{\text{feeding rate without mutation}} \quad . \quad (146)$$

The relative size of these classes in the nose h_0/h_1 will simply be ν_0/ν_1 since both classes will experience the same exponential growth while they establish. Previous work (33) has shown that the moment generating function for ν given a general feeding rate θ is simply,

$$\langle e^{-\theta z} \rangle \equiv e^{-\theta z^\alpha} \ , \quad (147)$$

where $\alpha \equiv 1 - s_b/x_{ca}$. More recent work (55) showed that $\rho_a(h|h_0)$ can be obtained via an inverse Laplace transform of ν_0/ν_1,

$$\rho_a(h|h_0) = \frac{\sin(\pi\alpha)}{\pi} \cdot \frac{\theta_0\theta_1}{h(1-h)\left[\theta_0^2 \cdot \left(\frac{1-h}{h}\right)^\alpha + \theta_1^2 \cdot \left(\frac{h}{1-h}\right)^\alpha + 2\theta_0\theta_1\cos(\pi\alpha)\right]} \ . \quad (148)$$

Assuming a log-uniform prior on h_0 and taking $h_0 \ll 1$, this work also showed that the unconditional distribution of mutation frequencies in the nose is,

$$\rho_a(h) = \frac{\sin(\pi\alpha)}{\pi} \cdot \frac{(1-h)^{\alpha-1}}{h^{1+\alpha}} \ , \quad (149)$$

where a mutation's probability of fixing is simply equal to h.

In future nose classes, mutations continue to compete to take over by producing linked descendants. As a result the conditional distribution of h in future nose classes $\rho(h, k|h_0)$ will vary with each subsequent establishment k. Kosheleva et al. showed that Eq. (148) can be generalized to obtain the present frequency spectrum of mutations that arose k classes in the past by substituting $\alpha \to \alpha^k$. After again enforcing the same assumptions on h_0, this yields an unconditional distribution,

$$\rho_a(h,k) = \frac{\sin(\pi\alpha^k)}{\pi} \cdot \frac{(1-h)^{\alpha^k-1}}{h^{1+\alpha^k}} \ . \quad (150)$$

Full frequency spectrum As the nose class grows to $\mathcal{O}(1)$ frequencies in the bulk, the distribution of mutations within the class will remain relatively constant. As a result, the full site frequency spectrum can be approximated from the site frequency spectrum in the bulk since most of the population resides there. To estimate this full site frequency spectrum, previous work (55) has shown we can sum over the contributions from increasingly old linked mutations $\rho(h,k)$,

$$\rho(f_a) \approx \sum_{k=0}^{\infty} \frac{\sin(\pi\alpha^k)}{\pi} \cdot \frac{(1-h)^{\alpha^k-1}}{h^{1+\alpha^k}} \ . \tag{151}$$

Since contributions from $k \gtrsim \mathcal{O}(x_{ca}/s_b)$ will fall off super-exponentially, this sum will be dominated by recently produced mutations. As we will show, it will be difficult to analytically calculate an analog of Eq. (148) in sexual populations. This makes it difficult to predict how the site frequency spectrum and genetic diversity evolve over time. To make progress, we will identify the key forces and processes influencing the fate of a new mutation in Eq. (148). This will allow us to develop crude approximations that can be used to study the dynamics in sexual populations.

9.3 Linked selection: jackpotting and busting

Previous work (35) has shown that Eq. (149) is a weighted average of two distinct clone frequency distributions. These contributions can be separated into a typical distribution, which occurs every $\sim \alpha$ establishments and an atypical distribution that occurs every $\sim 1-\alpha$ establishments. We explain the origin of these distributions below, first focusing on the typical distribution.

Typical distribution Provided the dynamics are typical, we can straightforwardly estimate the time τ_g to produce g successful mutations by integrating over the feeding rate,

$$\int_0^{\tau_g} x_{ca} \cdot U_b \cdot \frac{e^{(x_{ca}-s_b)t}}{x_{ca}} \, dt \ \approx g. \tag{152}$$

Integrating and solving for τ_g, we obtain,

$$\tau_g \approx \frac{1}{(x_{ca}-s_b)} \log\left(\frac{x_{ca}\cdot g}{U_b}\right) \ . \tag{153}$$

These individual mutations will each contribute to the establishment of the class as a whole. We can estimate this establishment time τ_q by summing over the growth of each individual mutation,

$$\tau_q \approx \frac{1}{(x_{ca}-s_b)} \log\left(\frac{s_b}{U_b}\right) \ , \tag{154}$$

where $q_a \equiv x_{ca}/s_b$. Once the whole class establishes, the g clones will begin to grow deterministically with an overall frequency $h_g \propto e^{x_{ca}(\tau_g-\tau_q)}$ that depends on the time they arose. This deterministic growth ensures that each mutation's relative size remains approximately frozen as it traverses to the bulk. This allows us to calculate the typical distribution of mutation sizes in the bulk as,

$$\frac{h_g}{h_1} \approx \frac{1}{g^{1/\alpha}} \ , \tag{155}$$

where h_1 is the frequency of the largest mutation in the nose. We can invert this rank distribution and take the derivative with respect to g to find the density of mutations at frequency h,

$$\rho_{typical}(h,1) = \begin{cases} \frac{1}{h^{1+\alpha}} \text{ if } h \lesssim 1/2 \ , \\ 0 \text{ if } h \gtrsim 1/2 \ , \end{cases} \quad , \tag{156}$$

where there are typically no mutations at a large frequency $h \gtrsim 1/2$. This result matches the scaling in Eq. (149).

Atypical distribution Rarely, a single mutation will establish atypically early in the nose and compose a majority of the nose class. The distribution of early establishment times τ_1 simply follows from the exponential feeding rate in Eq. (152),

$$\Pr(\tau_1) \equiv (x_{ca} - s_b) e^{(x_{ca}-s_b)(\tau_1 - \tau_q)} \tag{157}$$

The overall frequency of this early mutation will straightforwardly depend on the establishment time, $h \approx e^{x_{ca}(\tau_q-\tau_1)}/(1+e^{x_{ca}(\tau_q-\tau_1)})$. A standard transformation of this distribution shows that the large mutation's frequency will be distributed as,

$$\rho_{atypical}(h,1) \approx \alpha \cdot \frac{(1-h)^{\alpha-1}}{h^{1+\alpha}} \text{ if } h \gtrsim 1/2 \ . \tag{158}$$

This result demonstrates that lucky new mutations will typically only jump to a modest frequency $\int \rho_{atypical}(h,1) \cdot h \ dh \sim \log(2) + \frac{1}{2q_a}(1-\log(2))$. However, extremely large jumps will not be exceptionally uncommon, $\int_{h'}^{1} \rho_{atypical}(h,1) \sim (1-h')$.

Full mixed distribution The result in Eq. (150) is a mixture of these distributions,

$$\rho_a(h,1) \approx \rho_{typical}(h,1) + (1-\alpha) \cdot \rho_{atypical}(h,1) \ , \tag{159}$$

where the normalization reflects the overall probability of an early establishment. We can see this by integrating the total density of Eq. (150) above $1/2$, corresponding to the contribution from $\rho_{atypical}(h,1)$.

$$\int_{1/2}^{1} \frac{\sin(\pi\alpha)}{\pi} \cdot \frac{(1-h)^{\alpha-1}}{h^{1+\alpha}} dx = \frac{\sin(\pi\alpha)}{\pi\alpha} \approx (1-\alpha) = 1/q_a \tag{160}$$

This implies that an early establishment occurs roughly every $1/q_a$ establishments. These early establishments can strongly influence the fate of existing genetic variants in the population. This is because variants in the feeding class can attain a majority frequency by producing the early establishing mutation. Since the majority of variants will be at a small frequency, this process can lead to large frequency shifts in the typical successful variant. We refer to early establishments that allow linked small frequency mutations to jump to a majority frequency as *jackpot* events since they represent lucky jumps in frequency. Conversely, when an existing majority frequency mutation does not produce the early establishing mutation, it will drop down to a minority frequency. We refer to this process as *bust* events since they represent sudden unlucky drops in frequency. These dynamics motivate the approximation,

$$\rho_{atypical}(h|h_0) = \frac{(1-h)^{\alpha-1}}{h^{1+\alpha}} \text{ with } h_0 \gtrsim 1/2 \text{ with prob. } \frac{h_0}{q_a} \ . \tag{161}$$

9.4 Linked selection: genetic draft and genetic drowning

From the perspective of a competing mutation, jackpotting and busting represent atypical events. The kernel in Eq. (148) also shows that a mutation will typically experience modest directional changes in frequency every establishment. We can see this by defining $z_0 \equiv \text{logit}(h_0)$ and $z \equiv \text{logit}(h)$ and transforming the kernel in Eq. (148),

$$\rho_a(z|z_0) = \frac{\sin(\pi\alpha)/2\pi}{\cosh(\alpha z - z_0) + \cos(\pi\alpha)} \, . \tag{162}$$

This distribution corresponds to a Lévy distributed random walk in logit space that is symmetric around $z = z_0/\alpha$. This highlights that the actual diffusive dynamics will be biased depending on z_0. The expected step will depend on the initial frequency of the mutation $\int \rho_a(z|z_0)(z - z_0)\, dz = (\alpha - 1)z_0$. This indicates a threshold frequency near $z_0 \approx 1/2$, where mutations below $z_0 \lesssim 1/2$ will typically decline in frequency and mutations above $z_0 \gtrsim 1/2$ will typically increase in frequency. It also illustrates how jackpotting and busting, which shift variants across this threshold frequency, can have a compounding effect. These dynamics motivate a deterministic approximation,

$$\rho_{\text{typical}}(h, k|h_0) = \frac{(h_0)^{(1/\alpha)^k}}{(h_0)^{(1/\alpha)^k} + (1 - h_0)^{(1/\alpha)^k}} \approx \begin{cases} (h_0)^{(1/\alpha)^k} \text{ if } h_0 \lesssim 1/2 \, , \\ 1 - (1 - h_0)^{(1/\alpha)^k} \text{ if } h_0 \gtrsim 1/2 \, . \end{cases} \, , \tag{163}$$

where a mutation typically increases or decreases in frequency monotonically depending on its starting frequency. Following previous work (36)(55), we refer to the deterministic force that drives large-frequency mutations to further increase in frequency as *genetic draft*. We also deem the corollary deterministic force that purges low-frequency mutations as *genetic drowning*. As we will show, both of these processes will play an important role in sexual dynamics.

9.5 Trajectories of successful mutations

Since each mutation in the fitness class has the same relative fitness, its fixation probability will simply be proportional to its frequency. This allows to study the paths that successful mutations take by conditioning (or tilting) Eq. (162) on eventual fixation,

$$\rho_a(z|z_0, z_\infty = \infty) = \left(1 + e^{-z_0}\right) \cdot \frac{e^z}{1 + e^z} \cdot \frac{\sin(\pi\alpha^k)/2\pi}{\cosh(\alpha^k z - z_0) + \cos(\pi\alpha^k)} \, . \tag{164}$$

For sufficiently small frequency mutations, we can expand around $z_0 \to -\infty$ to show that,

$$\rho_a(z|z_0, z_\infty = \infty) \approx \frac{\sin(\pi\alpha^k)}{\pi} \cdot \frac{e^{-(\alpha^k - 1)z}}{1 + e^z} \, . \tag{165}$$

This expanded kernel is centered around $z^* \approx \log((1 - \alpha^k)/\alpha^k)$ suggesting that after $k \sim q_a \log(2)$ establishments a successful mutation will attain an appreciable frequency $z \sim 0$. Moreover in $k \sim q_a$ establishments, the mutation will only attain a modestly larger frequency. We can see this by taking the limit $x_{ca} \gg 1$, where z^* approaches $\log(e - 1) \sim 1/2$. This shows that jackpotting will represent the dominant force driving a successful mutation's change in frequency over the first q_a establishments. Conversely, for mutations already at a large frequency $z_0 \gtrsim 1/2$, we can expand Eq. (164) around $z \to \infty$ to show that the tilted operator will only weakly bias the dynamics. As a

result, the mutation will typically continue to experience deterministic growth similar to the approximation in Eq. (163). This shows that genetic draft will be the dominant force influencing the fate of successful large frequency mutations.

We can also similarly condition Eq. (162) on a mutation being unsuccessful. This allows us to show that failing mutations at a large frequency drop down to a minority within $k \sim q_a$ additional establishments, while failing mutations at small frequency continue to drown. These results allow us to crudely approximate the trajectory of new mutations after q_a establishments $\rho_a(h, k = q_a|h_0)$,

$$\rho_a(h, q_a|h_0) \approx \begin{cases} \mathcal{O}(1/2) \text{ with prob. } h_0 \text{ if } h_0 \lesssim 1/2 \ , \\ h_0^{1/\alpha^{q_a}} \text{ with prob. } 1-h_0 \text{ if } h_0 \lesssim 1/2 \ , \\ \mathcal{O}(1/2) \text{ with prob. } 1-h_0 \text{ if } h_0 \gtrsim 1/2 \ , \\ 1-(1-h_0)^{1/\alpha^{q_a}} \text{ with prob. } h_0 \text{ if } h_0 \lesssim 1/2 \ . \end{cases} \tag{166}$$

9.6 Sojourn times

These approximations have been used to reveal the timescale over which mutations persist and sweep in the population (55). In particular, the time for a mutation to transition from a single copy to a majority of the population will generally be limited by the rate of jackpotting. Jackpot events will occur at a rate $\sim 1/q_a$ establishments, allowing us to approximate,

$$\langle \tau_{1 \to N/2} \rangle \sim \sum_{n=1}^{\infty} \left(1 - \frac{1}{q_a}\right)^{n-1} \cdot \left(\frac{1}{q_a}\right) \cdot (q_a + n) \cdot \tau_{q,m} \sim \frac{2}{s_b} \cdot \log\left(\frac{s_b}{U_b}\right), \tag{167}$$

which accounts for the nose-to-mean timescale while the nose class grows to large overall frequency. Once a mutation attains this large frequency, it will need to fix by remaining linked to its future descendants. Its frequency will typically increase by a logit growth factor z_0/α^k over k subsequent establishments. This allows us to estimate the typical number of establishments to sweep in the nose as a function of a mutations initial logit frequency after jackpotting z_0,

$$k_{sw} \approx \frac{\log(\log N) - \log(z_0)}{\log(1/\alpha)} \ , \tag{168}$$

where the distribution of jackpot logit frequencies is $z_0 \propto e^{-\alpha z_0}$. Marginalizing over this distribution, allows us to estimate,

$$\langle k_{sw} \rangle \sim \frac{.577 + \log(\alpha \log(N))}{|\log \alpha|} \ . \tag{169}$$

This suggests the sojourn time from a majority frequency to fix is,

$$\langle \tau_{N/2 \to N} \rangle \approx \langle k_{sw} \rangle \cdot \tau_{q,m} \sim \frac{\log\log(N)}{s_b} \cdot \log\left(\frac{s_b}{U_b}\right) \ . \tag{170}$$

Provided $N \gg 10^4$, a mutation will spend the majority of its sweep time sojourning from a majority frequency to fixation. As we will show, this result will flip in sexual populations.

10 Recombination dominated dynamics: single transfer regime

We are now in a position to leverage these heuristic explanations to understand the genetic diversity and adaptive dynamics of sexual populations. We will continue to assume that most individuals in the population are concentrated near the mean fitness and that there is widespread clonal interference. This is equivalent to assuming,

$$x_c(t) - s_b \gg \sqrt{v(t)}, \tag{171}$$

$$s_b \gg \sqrt{v(t)}, \tag{172}$$

where we now allow for $x_c(t)$ and $v(t)$ to depend on the time-dependent genetic diversity. We also continue to assume that all transferred segments have a single genetic length $\tilde{\ell}$.

As we have discussed, the exact value of the recombination rate where quasi-asexuality breaks down will depend on parameters like x_{ca}/s_b and η. Moreover, the typical dynamics of successful mutations will also vary with these values. To understand the impact of genetic diversity on adaptive dynamics, we initially focus on the case where quasi-asexuality breaks down due to repeated transfers of single mutations. We refer to this regime as the *single transfer dominated regime*, which begins when $r \cdot \tilde{\ell} \gtrsim U_b$, Later, we show how this picture can be extended to the case where quasi-asexuality breaks down to repeated transfers of colocalized mutations. Provided,

$$\log\left(\frac{L}{\tilde{k}\cdot\tilde{\ell}}\right) \gtrsim \frac{3}{4}\cdot\log\left(NU_b\right) \text{ or } L\cdot f_b \lesssim \tilde{\ell}, \tag{173}$$

the typical mutation will arise distant from other mutations and take over by repeatedly acquiring single mutations from the bulk. We can understand these dynamics by studying how a fitness class arises in the nose and travels back to the bulk.

10.1 Stochastic dynamics In the nose

In a recombining population, the nose can be founded by new mutations and transferring existing mutations. Since transfer events confer a benefit of at most s_b, we can assume that the dominant acceptor background will be the same fitness class as the mutant feeding class. This allows us to write down a new linear branching process describing the founding of the nose,

$$\begin{aligned}\frac{\partial f(x_c,t)}{\partial t} =& \underbrace{x_c(t)\cdot f(x_c)}_{\text{selection}} + \underbrace{\frac{r\cdot\tilde{\ell}}{2}\cdot\sum_{x\le x_c-s_b}\sum_{i=0}^{f_b\cdot L}\left[h(g_i|x)\cdot(1-h(g_i|x_c-s_b))\right]\cdot f(x,t)\cdot f(x_c-s_b,t)}_{\text{recombination}} \\ &+ \underbrace{U_b\cdot f(x_c-s_b,t)}_{\text{mutation}} + \underbrace{\sqrt{\frac{f(x_c)}{N}}\eta(t)}_{\text{genetic drift}},\end{aligned} \tag{174}$$

where $h(g_i)$ are the frequencies of the $f_b\cdot L$ unique mutations within their respective fitness class. This nose class will be composed of a subclass of new mutations $f_m(x_c)$ and transferred mutations $f_r(x_c)$, where $f(x_c) = f_m(x_c) + f_r(x_c)$. These subclasses will compete with one another for

dominance with separate feeding rates. The recombinant class will obey the branching process,

$$\frac{\partial f_r(x_c,t)}{\partial t} = x_c(t)\cdot f_r(x_c) + \frac{r\cdot\tilde{\ell}}{2}\cdot \underbrace{\sum_{i=0}^{f_b\cdot L}[h(g_i|0,t)\cdot(1-h(g_i|x_c-s_b,t)]}_{\text{effective transfer rate } \omega(t)}\cdot f(x_c-s_b,t) + \sqrt{\frac{f_r(x_c)}{N}}\eta(t)\ , \tag{175}$$

where we have approximated the sum over the fitness classes with only the $x=0$ term. Since we have assumed $s_b \gg \sqrt{v(t)}$, the dominant contribution will originate from donors in the $x=0$ fitness class near the mean fitness, which is of frequency $f(x=0,t)\sim\mathcal{O}(1)$. We have also defined $\omega(t)$ as the effective transfer rate from the bulk to the nose. This rate will depend on the frequency of mutations in the donor class in the bulk as well as the acceptor class below the nose. The typical establishment time of the recombinant subclass in the nose will depend on this effective transfer rate $r\cdot\tilde{\ell}\cdot\omega(t)$,

$$\tau_{q,r} \approx \frac{1}{x_c(t)-s_b}\log\left(\frac{s_b}{r\cdot\tilde{\ell}\cdot\omega(t)}\right)\ . \tag{176}$$

While it is founded, the mutation founded subclass will continue to obey the linear branching process in Eq. (141), where the mean fitness now depends on recombination. This ensures it will establish over a similar time,

$$\tau_{q,m} \approx \frac{1}{x_c(t)-s_b}\log\left(\frac{s_b}{U_b}\right)\ . \tag{177}$$

10.2 Forward-time dynamical approach

The central challenge we face is that the patterns of ancestry that determine $\omega(t)$ and $x_c(t)$ will themselves depend on the dynamics in Eq. (174) and Eq. (175). This is because the relative size of each mutation in the bulk $h(g_i|0)$ will depend on the stochastic time at which it established in the nose $\sim x_c(t)/s_b$ mutations earlier. After establishing, the relative size of the clone within its fitness class will become effectively 'frozen' as it rises to large frequency in the bulk and has the greatest chance of being transferred. This creates a time-delayed feedback between the dynamics of Eq. (175) and the value of $\omega(t)$ roughly $x_c(t)/s_b$ establishments later. Moreover, a mutation's frequency in the feeding class $h(g_i|x_c-s_b)$ will implicitly depend on its frequency in the bulk $h(g_i|0)$. This is because a mutation can remain at large frequency by serving as an acceptor background during subsequent establishments.

To make progress, we can note that new genetic variation will not be injected into the population at a constant rate. When $\omega(t)$ is $\mathcal{O}(1)$, the nose will typically be dominated by transferred mutations since $r\cdot\ell \gg U_b$. We can see this by substituting the establishment times in Eq. (176) and Eq. (177) into the exponentially growing nose population in Eq. (143),

$$\left\langle\frac{f_r(x_c)}{f_m(x_c)}\right\rangle \approx e^{x_c(t)[\tau_{q,m}-\tau_q]} \sim \frac{r\cdot\tilde{\ell}\cdot\omega(t)}{U_b}\ . \tag{178}$$

As a result, new mutations will represent a small minority of the new nose and $\omega(t)$ will decrease in the future. Eventually if $\omega(t)$ becomes sufficiently small, then new mutations will reach a large enough frequency in the nose to replenish genetic diversity. Our central goal is to understand this process of

exhausting and replenishing genetic diversity. To begin, we will consider a simple initial condition, where the population is evolving asexually and recombination is suddenly turned on. We will then follow a single fitness class over multiple transfer cycles to understand how its genetic diversity declines with time. As we will show, this will lead to repetitive dynamics that constitute new steady state behavior.

10.3 Initial transfer rate

Provided the population has been adapting asexually, the frequency of mutations near the bulk $h(g_i|0) \equiv h_0$ and feeding class $h(g_i|x_c - s_b) \equiv h_{q-1}$ will initially resemble the asexual distribution. We can straightforwardly calculate the transfer rate from the asexual site frequency spectrum and transition kernel,

$$\omega(t_0) \approx \int_0^1 \int_0^1 \underbrace{h_0 \cdot (1-h_{q-1})}_{\text{sexual feeding rate}} \cdot \underbrace{\rho_a(h_0, 1)}_{\text{asexual SFS in bulk}} \cdot \underbrace{\rho_a(h_{q-1}, q-1|h_0)}_{\text{asexual linked selection}} \, dh_0 \, dh_{q-1} \, , \tag{179}$$

where we have neglected the contributions from transferring older linked mutations. Substituting Eq. (162) into Eq. (179), we obtain,

$$\omega(t_0) \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(e^{(1-\alpha)z_0 + z_{q-1}}) \cdot \sin(\pi\alpha) \cdot \sin(\pi\alpha^{q-1})}{2\pi^2 (1+e^{z_0}) \cdot (1+e^{z_{q-1}})^3 \cdot (\cosh(\alpha^{q-1} z_{q-1} - z_0) + \cos(\pi\alpha^{q-1}))} dz_0 \, dz_{q-1} \, , \tag{180}$$

where z_0 and z_{q-1} are the logit frequencies in the bulk and feeding class. While this integral is difficult to evaluate analytically, we can obtain an approximation using a simple scaling argument from our knowledge of the paths that mutations typically take. In particular we can approximate $\rho_a(h_{q-1}, q-1|h_0)$ using the kernel in Eq. (166). This allow us to approximate the dynamics in terms of whether a mutation in the donor class drafted, drowned, jackpotted, or busted in the acceptor class,

$$\begin{aligned} \omega(t_0) = \int_0^{\mathcal{O}(1/2)} [& \underbrace{h_0^2 \cdot \mathcal{O}(1/2)}_{\text{mutation jackpotted up}} + \underbrace{h_0 \cdot (1-h_0) \cdot (1 - h_0^{1/\alpha^{q-1}})}_{\text{mutation drowned down}}] \cdot \rho_a(h_0, 1) \, dh_0 \\ + \int_{\mathcal{O}(1/2)}^1 & [\underbrace{h_0^2 \cdot (1-h_0)^{1/\alpha^{q-1}}}_{\text{mutation drafted up}} + \underbrace{h_0 \cdot (1-h_0) \cdot \mathcal{O}(1/2)}_{\text{mutation busted down}}] \cdot \rho_a(h_0, 1) \, dh_0 \, , \end{aligned} \tag{181}$$

Substituting Eq. (150), we obtain,

$$\omega(t_0) \approx \alpha \, , \tag{182}$$

where the dominant contribution arises from mutations drowning down with $\mathcal{O}(1/q)$ contributions from mutations that busted down.

10.4 Fate of individual mutations

To predict how the site frequency spectrum and effective transfer rate change over time, we must follow the dynamics of each extant mutation g_j competing to establish in the nose. From the perspective of this focal mutation, it can jump to the nose by being a donor in the bulk, an acceptor in

the feeding class below the nose, or by producing a linked mutation in the feeding class,

$$\frac{dh(g_j|x_c,t)}{dt} = x_c(t)\cdot h_r(x_c) \; + \underbrace{U_b \cdot h(g_j|x_c - s_b, t)\cdot f(x_c - s_b, t)}_{\text{produced linked mutation}} + \sqrt{\frac{h_r(x_c)}{N}}\eta(t)$$
$$+ \frac{r\cdot\tilde{\ell}}{2}\cdot \underbrace{[h(g_j|0,t)\cdot(1-h(g_j|x_c-s_b,t))}_{\text{donated from bulk}} + \underbrace{(1-h(g_j|0,t))\cdot h(g_j|x_c-s_b,t)]}_{\text{accepted in feeding class}}\cdot f(x_c-s_b,t)\;. \tag{183}$$

We can straightforwardly see that this branching process takes the same form as the asexual one in Eq. (144) with a distinct non-linear feeding rate,

$$\theta_0 = \frac{r\cdot\tilde{\ell}}{2}\cdot[h_0\cdot(1-h_{q-1}) + (1-h_0)\cdot h_{q-1}]\;, \tag{184}$$

where we have neglected contributions from mutation and again employed the simplified notation $h(g_j|0)\equiv h_0$ and $h(g_j|x_c-s_b)\equiv h_{q-1}$. As the mutation establishes in the nose, it will compete against a subpopulation feeding recombinants without the mutation at rate θ_1,

$$\theta_1 = \frac{r\cdot\tilde{\ell}}{2}\cdot[(1-h_0)\cdot(1-h_{q-1})]\;. \tag{185}$$

From this setup, we can already see that genetic draft and genetic drowning will operate over separate timescales in sexual populations. Once a mutation attains a majority in its fitness class, it can quickly grow to become the dominant acceptor background by remaining linked to future transferring mutations. This growth is analogous to linked selection in asexual populations and occurs over short timescales (e.g. the time for a nose class to reach the bulk). Conversely, mutations in sexual populations can remain extant for much longer than in asexual populations. This is because a new nose class must wait $x_c(t)/s_b$ subsequent establishments until it reaches the bulk again and becomes the dominant donor class. This suggests that genetic drowning will drive mutations to extinction over much longer timescales (e.g. $x_c(t)/s_b$ multiples of the nose to mean timescale).

10.5 First update of the site frequency spectrum

To understand how these processes influence the transfer rate over time, we will need to follow the site frequency spectrum over multiple nose-to-mean timescales. This will require us to make several important generalizations from the asexual case. First, since the width of the fitness distribution $x_c(t)$ is now time-dependent, α can change from its initial value. To account for this time-dependence, we will need to generalize the transition kernel to depend on $\alpha_\gamma \equiv 1 - s_b/x_c(\gamma(t))$ where γ indexes the number of nose-to-mean transfer cycles since t_0. Secondly, since the feeding rates in Eq. (184) and Eq. (185) depend on h_0 and h_{q-1}, where h_{q-1} is a function of h_0, the transition kernel will have an implicit dependence. These considerations require us to define a generalized transition kernel $\rho_a(h|h_0) \to \rho_r(h|h_0, h_{q-1}, \gamma)$ that depends on the feeding rates in Eq. (184) and Eq. (185) and site frequency spectrum $\rho_r(h,\gamma)$. Crucially, this new transition kernel represents the change in frequency over a nose-to-mean timescale versus a single establishment time in asexual populations. To capture the effect of linked selection from the bulk to the acceptor class, we can also extend the asexual kernel

to depend on α_γ, $\rho_a(h,k|h_0) \to \rho_a(h,k|h_0,\gamma)$, where k continues to index the number of establishments. The site frequency spectrum after the first update is then simply,

$$\rho_r(h,0) \equiv \lim_{h'\to 0} h^{-1} \cdot \rho_r(h|h',0) \;, \tag{186}$$

where $\gamma = 0$ indexes the first transfer and we apply the same log-uniform assumption on h' used to derive Eq. (149). The transition kernel is given by,

$$\rho_r(h|h',0) \equiv \int_0^1 \int_0^1 \underbrace{\rho_r(h|h_0,h_{q-1},0)}_{\text{new sexual nose SFS}} \cdot \underbrace{\rho_a(h_{q-1},q-1|h_0,0)}_{\text{linked selection in acceptor class}} \cdot \underbrace{\rho_a(h_0|h')}_{\text{initial asexual SFS}} \; dh_{q-1}dh_0 \;. \tag{187}$$

Following the same methods in previous work (55), we can show that these integrals can be further simplified,

$$\rho(\tilde{h}|h',0) = \\ \mathrm{Re}\left[\int_0^\infty \frac{e^{-z^{\alpha_0\alpha_1}\left[h_0 e^{i\alpha_0\alpha_1\frac{\pi}{2}}+(1-h_0)\tilde{h}^{\alpha_0\alpha_1}e^{-i\alpha_0\alpha_1\frac{\pi}{2}}\right]-z^{\alpha_0\alpha_1^{q-1}}h_0^{q-1}\left[(1-2h_0)e^{i\alpha_0\alpha_1^{q-1}\alpha_0\frac{\pi}{2}}+(1-h_0)\tilde{h}^{\alpha_0\alpha_1^{q-1}}e^{-i\alpha_0\alpha_1^{q-1}\frac{\pi}{2}}\right]}}{\pi\cdot i\left[-(1-h_0)\alpha_0\alpha_1(\tilde{h}z)^{\alpha_0\alpha_1-1}e^{-i(\alpha_0\alpha_1-1)\frac{\pi}{2}}-(1-h_0)h_0^{q-1}\alpha_0\alpha_1^{q-1}(\tilde{h}z)^{\alpha_0\alpha_1^{q-1}-1}e^{-i(\alpha_0\alpha_1^{q-1}-1)\frac{\pi}{2}}\right]^{-1}} dz\right], \tag{188}$$

where $\tilde{h} \equiv h/(1+h)$ is the relative frequency of the mutation. While this integral does not have a simple closed form because of the implicit dependence of h_{q-1} on h_0, we can again leverage our understanding of mutation's typical paths to approximate $\rho_a(h_{q-1},q-1|h_0,0)$. This allows us to coarse-grain,

$$\begin{aligned}
\rho_r(h|h',0) \approx & \underbrace{\int_0^{\mathcal{O}(1/2)} \rho_a(h_0|h')\cdot h_0 \cdot \rho_{r,s,j}(h|h_0,\mathcal{O}(1/2),0)\; dh_0}_{\text{mutation small in bulk and jackpotted up in acceptor class}} \\
& + \underbrace{\int_0^{\mathcal{O}(1/2)} \rho_a(h_0|h')\cdot (1-h_0) \cdot \rho_{r,s,d}(h|h_0,h_0^{1/\alpha_0^{q-1}},0)\; dh_0}_{\text{mutation small in bulk and drowned down in acceptor class}} \\
& + \underbrace{\int_{\mathcal{O}(1/2)}^1 \rho_a(h_0|h')\cdot h_0' \cdot \rho_{r,l,d}(h|h_0,(1-(1-h_0)^{1/\alpha_0^{q-1}}),0)\; dh_0}_{\text{mutation large in bulk and drafted up in acceptor class}} \\
& + \underbrace{\int_{\mathcal{O}(1/2)}^1 \rho_a(h_0|h')\cdot (1-h_0') \cdot \rho_{r,l,b}(h|h_0,\mathcal{O}(1/2),0)\; dh_0}_{\text{mutation large in bulk and busted down in acceptor class}} \;.
\end{aligned} \tag{189}$$

Now that we have removed the implicit dependence on h_{q-1}, these integrals can be approximated provided we define the feeding rates $\theta_0(h_0,h_{q-1})$ and $\theta_1(h_0,h_{q-1})$ for $\rho_r(h|h_0,h_{q-1},0)$. We consider these contributions separately below.

Mutations initially small that jackpotted up Provided a mutation jackpotted up in the acceptor class, the feeding rates in Eq. (184) and Eq. (185) can be simplified,

$$\theta_0 \approx \frac{r \cdot \tilde{\ell}}{2}[\underbrace{h_0 \cdot \mathcal{O}(1/2)}_{\text{mutation donated from bulk}} + \underbrace{(1-h_0) \cdot \mathcal{O}(1/2)}_{\text{mutation acceptor in feeding class}}] , \tag{190}$$

$$\theta_1 \approx \frac{r \cdot \tilde{\ell}}{2} \underbrace{(1-h_0) \cdot \mathcal{O}(1/2)}_{\text{mutation not sampled}} . \tag{191}$$

Provided the mutation is initially small, the second term in Eq. (190) will dominate, ensuring the feeding rates are roughly equal. In this case the transition kernel becomes independent of h_0 and we obtain,

$$\rho_{r,s,j}(h,0) \approx \frac{(1-\alpha_0)}{2} \cdot \frac{\sin(\pi\alpha_1)}{\pi} \cdot \frac{1}{h(1-h)\left[\left(\frac{1-h}{h}\right)^{\alpha_1} + \cdot \left(\frac{h}{1-h}\right)^{\alpha_1} + 2\cos(\pi\alpha_1)\right]} . \tag{192}$$

Mutations initially small that drowned down Provided a mutation was small in the bulk and drowns in the acceptor class, we can express Eq. (184) and Eq. (185) as,

$$\theta_0 \approx \frac{r \cdot \tilde{\ell}}{2}[\underbrace{h_0 \cdot (1 - h_0^{1/\alpha_0^{q-1}})}_{\text{mutation donated from bulk}} + \underbrace{(1-h_0) \cdot h_0^{1/\alpha_0^{q-1}}}_{\text{mutation acceptor in feeding class}}] , \tag{193}$$

$$\theta_1 \approx \frac{r \cdot \tilde{\ell}}{2} \cdot \underbrace{(1-h_0) \cdot (1-h_0^e)}_{\text{mutation not sampled}} . \tag{194}$$

Since we assume h_0 is small, Eq. (193) will be dominated by contributions from transferring from the bulk. This leads to feeding rates identical to the asexual kernel and allows us to approximate,

$$\rho_{r,s,d}(h|h',0) \approx \int_0^{\mathcal{O}(1/2)} (1-h_0) \cdot \rho_a(h|h_0,0) \cdot \rho_a(h_0|h') \, dh_0 . \tag{195}$$

Since we are focused on small frequency mutations, we can approximate the $(1-h_0)$ factor as $\mathcal{O}(1)$. This implies that contributions to the site frequency spectrum will take the same form as the asexual site frequency spectrum with an effective $\alpha \to \alpha_0 \cdot \alpha_1$,

$$\rho_{r,s,d}(h,0) \approx \frac{\sin(\pi\alpha_1\alpha_2)}{\pi} \cdot \frac{(1-h)^{\alpha_0\alpha_1 - 1}}{h^{1+\alpha_0\alpha_1}} . \tag{196}$$

Mutations initially large that draft up Provided a mutations drafts upward, the feeding rates in Eq. (184) and Eq. (185) can be approximated as,

$$\theta_0 \approx \frac{r \cdot \tilde{\ell}}{2}[\underbrace{h_0 \cdot (1-h_0)^{1/\alpha_0^{q-1}}}_{\text{mutation donated from bulk}} + \underbrace{(1-h_0) \cdot (1-(1-h_0)^{1/\alpha_0^{q-1}})}_{\text{mutation acceptor in feeding class}}] . \tag{197}$$

$$\theta_1 \approx \frac{r \cdot \tilde{\ell}}{2} \underbrace{(1-h_0) \cdot (1-h_0)^{1/\alpha_0^{q-1}}}_{\text{mutation not sampled}} . \tag{198}$$

Since the mutation was initially large $h_0 \gtrsim 1/2$, the second term in Eq. (197) will dominate and, after factoring out shared terms, we obtain that $\theta_0 \sim 1$ and $\theta_1 \sim (1-h_0)^{1/\alpha_0^{q-1}}$. This leads to the kernel,

$$\rho_{r,l,d}(h|h_0) \approx \frac{\sin(\pi\alpha_1)/\pi}{h(1-h)\left[(1-h_0)^{\alpha_0^{q-1}}\left(\frac{1-h}{h}\right)^{\alpha_1} + (1-h_0)^{1/\alpha_0^{q-1}} \cdot \left(\frac{h}{1-h}\right)^{\alpha_1} + 2\cdot\cos(\pi\alpha_1)\right]} \,. \tag{199}$$

To leading order in $h_0 \gtrsim 1/2$, this allows us to approximate,

$$\rho_{r,l,d}(h,0) \approx (1-\alpha_0)\cdot\frac{\sin(\pi\alpha_0^q\alpha_1)}{\pi}\frac{(1-h)^{\alpha_0^q\alpha_1-1}}{h^{\alpha_0^q\alpha_1+1}} \,. \tag{200}$$

Mutations initially large that bust down Provided a mutation busts down we can approximate,

$$\theta_0 \approx \frac{r\cdot\tilde{\ell}}{2}[\quad \underbrace{h_0\cdot\mathcal{O}(1/2)}_{\text{mutation donated from bulk}} \quad + \quad \underbrace{(1-h_0)\cdot\mathcal{O}(1/2)}_{\text{mutation acceptor in feeding class}} \quad] \,. \tag{201}$$

$$\theta_1 \approx \frac{r\cdot\tilde{\ell}}{2}\underbrace{(1-h_0)\cdot\mathcal{O}(1/2)}_{\text{mutation not sampled}} \,. \tag{202}$$

Provided $h_0 \gtrsim 1/2$, the first term will dominate in Eq. (201) and the feeding rates will be roughly equal. This again allows to approximate,

$$\rho_{r,l,b}(h,0) \approx \frac{(1-\alpha_0)}{2}\cdot\frac{\sin(\pi\alpha_1)}{\pi}\cdot\frac{1}{h(1-h)\left[\left(\frac{1-h}{h}\right)^{\alpha_1} + \left(\frac{h}{1-h}\right)^{\alpha_1} + 2\cos(\pi\alpha_1)\right]} \,. \tag{203}$$

10.6 First update to the transfer rate

We are now in a position to understand how the new coarse-grained site frequency spectrum influences the next transfer rate when $\gamma = 1$.

$$\omega(1) \approx \int_0^1\int_0^1 h_0\cdot(1-h_{q-1})\cdot[\rho_{r,s,j}(h_0)+\rho_{r,l,d}(h_0)+\rho_{r,s,d}(h_0)+\rho_{r,l,b}(h_0)]\cdot\rho_a(h_{q-1},q-1|h_0,1)\,dh_0\,dh_{q-1} \,. \tag{204}$$

We can again approximate the dynamics of $\rho_a(h_{q-1},q-1|h_0,1)$ using the kernel in Eq. (166).

$$\begin{aligned}\omega(1) \approx \int_0^{\mathcal{O}(1/2)} [\quad & \underbrace{h_0^2\cdot\mathcal{O}(1/2)}_{\text{mutation jackpotted up}} + \underbrace{h_0\cdot(1-h_0)\cdot(1-h_0^{1/\alpha_1^{q-1}})}_{\text{mutation drowned down}}]\cdot[\rho_{r,s,j}(h_0)+\rho_{r,l,d}(h_0)+\rho_{r,s,d}(h_0)+\rho_{r,l,b}(h_0)]\;dh_0 \\ + \int_{\mathcal{O}(1/2)}^1 & [\underbrace{h_0^2\cdot(1-h_0)^{1/\alpha_1^{q-1}}}_{\text{mutation drafted up}} + \underbrace{h_0\cdot(1-h_0)\cdot\mathcal{O}(1/2)}_{\text{mutation busted down}}]\cdot[\rho_{r,s,j}(h_0)+\rho_{r,l,d}(h_0)+\rho_{r,s,d}(h_0)+\rho_{r,l,b}(h_0)]\;dh_0 \,.\end{aligned} \tag{205}$$

Substituting our solutions above, we can see this integral will be dominated by contributions from $\rho_{r,s,d}(h_0)$, corresponding to mutations at small frequency in the bulk that drowned down in the acceptor class. This allows us to approximate,

$$\omega(1) \approx \frac{\sin(\pi\alpha_0\alpha_1)}{\pi}\int_0^{\mathcal{O}(1/2)}\frac{(1-h_0)^{\alpha_0\alpha_1}}{h_0^{\alpha_0\alpha_1}}dh_0 \approx \frac{\sin(\pi\alpha_0\alpha_1)}{\pi}\cdot\mathcal{O}(1/2) \tag{206}$$

10.7 General updates of the transfer rate and site frequency spectrum

We can straightforwardly iterate these dynamics forward in time to calculate the site frequency spectrum and transfer rate. As we do so, the dominant contribution to the transfer rate will come from small frequency mutations that drown down over multiple nose-to-mean timescales. This allows us to show that the transfer rate will typically fall off as,

$$\omega(\gamma) \approx \prod_{i=0}^{\gamma} \alpha_i \ , \tag{207}$$

where we have expanded around $\omega(\gamma) \ll 1$. While this value will self-consistently depend on how $x_c(t)$ varies with each establishment, we can show that it will decline relatively slowly compared to the timescale over which the nose establishes. In particular, over one nose-to-mean timescale, involving $q(0) \gg 1$ establishments, the transfer rate will only decline by a factor of $1 - 1/q(0) \sim 1$. This motivates a simple estimate of q based on asexual evolution with an effective mutation rate $U_b \to r \cdot \tilde{\ell} \cdot \omega(\gamma)$,

$$q(\gamma) \approx \frac{2\log(Ns_b)}{\log(s_b/r \cdot \tilde{\ell} \cdot \omega(\gamma))} \approx \frac{2\log(Ns_b)}{\log(s_b/r \cdot \tilde{\ell}) - \sum_{\gamma'=0}^{\gamma} \log(1 - \frac{1}{q(\gamma')})} \ , \tag{208}$$

where we have substituted for $\alpha_\gamma = 1 - 1/q(\gamma)$. Together, Eq. (208) and Eq. (207), yield an exact ordinary differential equation solution for $q(\gamma)$ in the limit where γ is treated as a continuous variable,

$$\frac{dq(\gamma)}{d\gamma} = q(\gamma) \cdot \frac{\log(1 - \frac{1}{q(\gamma)})}{\log(s_b/r \cdot \tilde{\ell})} \ . \tag{209}$$

While this can be solved exactly with special functions, we can obtain accurate approximations assuming $q(\gamma) \gg 1$ for the relevant γ. This allows us to expand Eq. (209) to second order and solve,

$$q(\gamma) \approx \left(q(0) + \frac{1}{2}\right) \cdot e^{-\frac{\gamma}{2\log(Ns_b)}} - \frac{1}{2} \ . \tag{210}$$

Similarly, the transfer rate obeys the differential equation,

$$\frac{d\omega}{d\gamma} = \omega(\gamma) \log\left(1 - \frac{1}{q(0)} + \frac{\log(\omega(\gamma))}{2\log(Ns_b)}\right) \ . \tag{211}$$

A second order solution in $q(\gamma) \gg 1$ yields,

$$\omega(\gamma) \approx e^{-\frac{\gamma}{q(0)} - \frac{\gamma^2}{2q(0)^2 \log(s_b/r \cdot \tilde{\ell})} - \frac{\gamma}{2q(0)^2}} , \tag{212}$$

where we have also expanded to second order around $\gamma \ll 2\log(Ns_b)$ This shows that the transfer rate will initially fall off exponentially fast. We are now in a position to ask what values of γ are attained before new mutations can replenish genetic diversity.

10.8 Depletion and exhaustion of genetic diversity

Thus far, we have calculated how the transfer rate declines with each nose-to-mean timescale γ. To do this, we have assumed that the dominant contribution to the transfer rate comes from the original class

of asexual mutations. However, with each transfer cycle, a new subclass of mutations will be introduced into the focal fitness class. These new mutations will also contribute to the transfer rate relative to their overall frequency and age. We are now in a position to consider these contributions and evaluate our assumption that the original class remains dominant. To do this, we define a general transfer rate at transfer cycle γ,

$$\omega_{total}(\gamma) \equiv \sum_{i=0}^{\gamma} f_{i,\gamma} \cdot \omega(\gamma - i) \ , \tag{213}$$

where $f_{i,\gamma}$ denotes the frequency of mutations in the bulk at transfer cycle γ produced during transfer cycle i. Since our solution for $\omega(\gamma - i)$ in Eq. (207) is dominated by small frequency mutations, we can generalize it to describe aging in new subdominant mutation classes. This allows us to simply calculate the total transfer rate forward in time accounting for the contributions of each new subclass of mutations. We can then ask: at what value of γ do new mutations begin to attain high frequencies in the nose?

During the first transfer after recombination is turned on, all mutations will have been produced $q(0)$ establishments earlier. The transfer rate will therefore simply be,

$$\omega_{total}(0) = \underbrace{r \cdot \tilde{\ell} \cdot \alpha_0 \cdot f_{0,0}}_{\text{transfer rate from bulk}} \ , \tag{214}$$

where we have substituted $\omega(0) = \alpha_0$ and by definition $f_{0,0} \sim 1$. As these mutations transfer to the nose, they will compete with a subclass of new mutations with feeding rate U_b. The typical ratio of old versus new mutations will depend on these feeding rates and genetic draft, which amplifies the larger class of mutations,

$$\frac{f_{1,1}}{f_{0,1}} \sim \left(\frac{U_b}{r \cdot \tilde{\ell} \cdot \alpha_0} \right)^{\frac{1}{\alpha_0}} \ . \tag{215}$$

Once the new nose reaches the bulk, the next transfer cycle $\gamma = 1$ will have mutations of two different ages contributing to the transfer rate,

$$\omega_{total}(1) = f_{0,1} \cdot \alpha_0 \cdot \alpha_1 + f_{1,1} \cdot \alpha_1 \tag{216}$$

Provided $\log(r \cdot \tilde{\ell}/U_b) \gtrsim 2/q(0)$, the transfer rate will be dominated by the original class of mutations. In the next nose, we can approximate their relative sizes,

$$\frac{f_{2,2}}{f_{0,2}} \approx \left(\frac{U_b}{r \cdot \tilde{\ell} \cdot f_{0,1} \cdot \alpha_0 \alpha_1} \right)^{\frac{1}{\alpha_1}} \sim \left(\frac{U_b}{r \cdot \tilde{\ell} \cdot \alpha_0 \alpha_1} \right)^{\frac{1}{\alpha_1}} \ , \tag{217}$$

$$\frac{f_{1,2}}{f_{0,2}} \approx \left(\frac{f_{1,1}}{f_{0,1} \cdot \alpha_0} \right)^{\frac{1}{\alpha_1}} \ . \tag{218}$$

As these classes grow to a large overall frequency, they will influence the $\gamma = 2$ transfer rate,

$$\omega_{total}(2) \ \approx f_{0,2} \cdot \alpha_0 \cdot \alpha_1 \cdot \alpha_2 \ + f_{1,2} \cdot \alpha_1 \cdot \alpha_2 \ + f_{2,2} \cdot \alpha_2 \ . \tag{219}$$

These dynamics lead to a general recursive solution for $\omega_{total}(\gamma)$,

$$\omega_{total}(\gamma) \approx f_{0,\gamma} \cdot \prod_{i=0}^{\gamma} \alpha_i \cdot \left[1 + \sum_{j=1}^{\gamma} \left(\frac{r \cdot \tilde{\ell}}{U_b} \right)^{-\prod_{i=j-1}^{\gamma-1} \frac{1}{\alpha_i}} \cdot \left(\prod_{i=0}^{j-1} \frac{1}{\alpha_i} \right)^{\sum_{t=j-1}^{\gamma-1} \prod_{m=t}^{\gamma-1} \frac{1}{\alpha_m}} \cdot \left(\prod_{i=0}^{j-1} \frac{1}{\alpha_i} \right) \right] \ , \tag{220}$$

which can be expressed as,

$$\omega_{total}(\gamma) \approx f_{0,\gamma} \cdot \omega(\gamma) \cdot \left[1 + \sum_{j=1}^{\gamma} \left(\frac{r \cdot \tilde{\ell}}{U_b} \right)^{-\frac{\omega(j-2)}{\omega(\gamma-1)}} \cdot \omega(j-1)^{-\sum_{t=j-1}^{\gamma-1} \frac{\omega(t-1)}{\omega(\gamma-1)}} \cdot \omega(j-1)^{-1} \right] , \quad (221)$$

Now since $q_a \gg 1$, we can approximate,

$$\omega_{total}(\gamma) \propto 1 + \sum_{j=1}^{\gamma} e^{\frac{\gamma-j+1}{q_0} \left[j - \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) \right] - j(1-\frac{1}{q_0})} . \quad (222)$$

The original class of asexual mutations will be overtaken when the contributions from the sum become larger than 1. This will correspond to a critical value of γ that we refer to as γ^*. Similarly, the sum will typically be maximized at a characteristic value of j that we refer to as j^*. This value will obey the critical point equation,

$$\gamma^* - j^* + 1 = q(0) \cdot \log \left(\frac{q(0) - 1}{q(0) - j^* + \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)} \right) . \quad (223)$$

We can also show that at this critical point, terms within,

$$\Delta j \sim \sqrt{q(0)} , \quad (224)$$

will contribute to the sum. This allows us to evaluate the sum with a Gaussian Laplace approximation and estimate γ^*,

$$\gamma^* - j^* + 1 = q(0) \log \left(\frac{j^* \cdot (1 - \frac{1}{q(0)}) - \frac{1}{2} \log(q(0))}{j^* - \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)} \right) , \quad (225)$$

By setting Eq. (225) equal to Eq. (223) we can solve for j^*,

$$j^* = \frac{\big(q(0)-1\big) \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) + \frac{q(0)}{2} \log q(0)}{2\big(q(0)-1\big)} + \quad (226)$$

$$\frac{\sqrt{\left(\big(q(0)-1\big) \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) + \frac{q(0)}{2} \log q(0)\right)^2 - 4\big(q(0)-1\big) \left(\frac{q(0)}{2} \left(q(0) + \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) \right) \log q(0) - q(0)\big(q(0)-1\big) \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) \right)}}{2\big(q(0)-1\big)} . \quad (227)$$

To first order in $q(0) \gg 1$, we obtain the solution for j^*,

$$j^* = \frac{1}{2} \log \left(\frac{r \cdot \tilde{\ell}}{U_b} \right) + \frac{1}{2} \sqrt{\log^2 \left(\frac{r \cdot \tilde{\ell}}{U_b} \right) + 4q(0) \cdot \log \left(\frac{r \cdot \tilde{\ell}}{U_b} \right)} \sim \sqrt{q(0) \cdot \log \left(\frac{r \cdot \tilde{\ell}}{U_b} \right)} , \quad (228)$$

where we have expanded around $4 \cdot q(0) \gg \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)$ on the right hand side. This solution can be substituted into Eq. (225) to show that the original class of asexual mutations will become

subdominant when,

$$\gamma^* \sim 2 \cdot \sqrt{q(0) \cdot \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)} + \frac{1}{2}\log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) - 2 - \frac{\log(q(0))}{2} \cdot \sqrt{\frac{q(0)}{\log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)}} \sim 2 \cdot \sqrt{q(0) \cdot \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)}\,, \tag{229}$$

where the first term dominates and we have expanded to first order in $4q(0) \gtrsim \log(r \cdot \tilde{\ell}/U_b)$. This allows us to show that our assumptions that $q(\gamma) \gg 1$ and $\gamma \ll 2\log(Ns_b)$ are valid provided $\log(r \cdot \tilde{\ell}/U_b) \ll \log(Ns_b) \cdot \log(s_b/r \cdot \tilde{\ell})$, which is true since $q_a \gg 1$.

At the critical value of γ^*, we can also show that the transfer rate will bottom out at,

$$\omega(\gamma^*) \sim e^{-2\sqrt{\frac{\log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)}{q(0)}}}\,, \tag{230}$$

with a corresponding width of the fitness distribution,

$$q(\gamma^*) \sim q(0) \cdot \left(1 - \sqrt{\frac{2\log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)}{\log\left(\frac{s_b}{r \cdot \tilde{\ell}}\right) \cdot \log\left(Ns_b\right)}}\right)\,. \tag{231}$$

These two values are important because they characterize the magnitude of the oscillations in fitness and genetic diversity that occur in sexual populations. In particular, they show that the magnitude of oscillations will increase with $r \cdot \tilde{\ell}$, but decrease with the population size.

10.9 Replenishment of genetic diversity and steady state dynamics

As the transfer rate bottoms out, a privileged window for new genetic variants to dominate the transfer rate will occur. We can see this by asking which mutations dominate the immediately next transfer rate. Specifically, the next transfer rate will be,

$$\omega_{total}(\gamma^* + 1) \propto \sum_{j=j^*}^{\gamma^*} e^{e^{\frac{\gamma^* - j + 1}{q_0}\left[j - \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right)\right] - \frac{j}{q_0}}}\,. \tag{232}$$

This is a strictly increasing function with respect to j, so that contributions from γ^* will dominate the next transfer. This implies the transfer rate will revert back to $\mathcal{O}(1)$, reproducing the initial asexual steady state. Interestingly, this shows that genetic diversity will be replenished even while newly produced mutations never reach a large frequency in the population overall.

10.10 Upper bound on the period

Thus far, in our analysis of the period of oscillations we have made a crucial assumption that the recycled genetic variation driving the transfer rate does not go extinct. In reality, as the value of γ^* grows increasingly large, mutations in the original burst will eventually go extinct before the predicted value in Eq. (261) is attained. This process will depend on the population size and sets an upper bound on the value of γ^* We can crudely estimate this value by focusing on the largest mutations in the founding burst. These mutations will initially attain a modest overall frequency $1\% \lesssim f \lesssim 50\%$ and

then typically decline with each transfer cycle. As a result, we expect this typical mutation to go extinct in a number of generations determined by genetic drowning,

$$-\log(N) \sim \log(f_0) \cdot \prod_{i=0}^{\gamma^*_{\max}} \alpha_i \ . \tag{233}$$

This allows us to approximate,

$$\gamma^*_{\max} \sim q(0) \log(\log(N)) \ . \tag{234}$$

10.11 Long-term estimates of the rate of adaptation

We can use these results to obtain long-term average estimates for the rate of adaptation. Specifically, we can sum over the total increase in fitness while the population runs out of mutations,

$$v \approx \frac{\sum_{\gamma=0}^{\gamma^*} q(\gamma) \cdot s_b}{\sum_{\gamma=0}^{\gamma^*} \tau_{q,r}(\gamma) \cdot q(\gamma)} \ . \tag{235}$$

Substituting the value for $\tau_{q,r}(k)$ in Eq. (176) and $q(k)$ in Eq. (208) we obtain,

$$\langle v(t) \rangle \approx \frac{2 s_b^2 \log(N\sqrt{s_b \cdot r \cdot \tilde{\ell}})}{\log(s_b / r \cdot \tilde{\ell})} \cdot \left[1 - \frac{\gamma^* + 1}{4 \log(N s_b)}\right]^2 \ . \tag{236}$$

Substituting our approximation for γ^*, we can approximate

$$\langle v(t) \sim \frac{2 s_b^2 \log(N\sqrt{r \cdot \tilde{\ell} s_b})}{\log(s_b / r \cdot \tilde{\ell})} \cdot \left[1 - \sqrt{\frac{\log(r \cdot \tilde{\ell}/U_b)}{2 \cdot \log(N s_b) \cdot \log(s_b / r \cdot \tilde{\ell})}}\right]^2 \ . \tag{237}$$

We can similarly obtain a long-time average estimate of q,

$$\langle q(t) \rangle \approx \frac{1}{k^*} \sum_{k=0}^{k^*} \frac{2 \log(N s_b) - k}{\log(s_b / r \cdot \tilde{\ell})} \approx q(0) \cdot \left[1 - \sqrt{\frac{\log(r \cdot \tilde{\ell}/U_b)}{2 \cdot \log(N s_b) \cdot \log(s_b / r \cdot \tilde{\ell})}}\right] \tag{238}$$

10.12 Duration of period

We are now in a position to estimate the duration of a period,

$$T_{per} = \sum_{k=0}^{\gamma^*} q(k) \cdot \tau_{q,r}(k) \ . \tag{239}$$

To first order we obtain,

$$T_{per} \approx \frac{\gamma^*}{s_b} \cdot \left(1 + \frac{1}{q_0}\right) \cdot \left[\log\left(\frac{s_b}{r \cdot \tilde{\ell}}\right) + \frac{\gamma^*}{2 q_0}\right] \sim \frac{2}{s_b} \sqrt{2 \cdot \log(N s_b) \cdot \log\left(\frac{r \cdot \tilde{\ell}}{U_b}\right) \cdot \log\left(s_b / r \cdot \tilde{\ell}\right)} \ . \tag{240}$$

This solution allows us to estimate the sojourn and fixation times of successful mutations.

10.13 Sojourn times

Our analysis above shows that recombination will influence the trajectory of recombined mutations in distinct and counteracting ways. On the one hand, the sojourn time to grow from one copy to a majority will increase as $r \cdot \tilde{\ell} \to U_b$. We can immediately see this from our understanding of quasi-asexual dynamics. Successful recombined mutations will need to transfer once before attaining a large frequency. This extends the average sojourn time by a factor proportional the fixation rate of recombined mutations,

$$\langle \tau_{1\to N/2} \rangle \sim \left(\frac{1}{1+\frac{r\cdot\tilde{\ell}}{U_b}} \right) \cdot \sum_{n=1}^{\infty} (q_a+n) \cdot \left(1-\frac{1}{q_a}\right)^{n-1} \cdot \left(\frac{1}{q_a}\right) \cdot \tau_{q,m} + \left(\frac{\frac{r\cdot\tilde{\ell}}{U_b}}{1+\frac{r\cdot\tilde{\ell}}{U_b}} \right) \sum_{n=1}^{\infty} (2q_a+n) \cdot \left(1-\frac{1}{q_a}\right)^{n-1} \cdot \left(\frac{1}{q_a}\right) \cdot \tau_{q,m} \, . \tag{241}$$

This can be evaluated,

$$\langle \tau_{1\to N/2} \rangle \sim \frac{2+3\cdot\frac{r\cdot\tilde{\ell}}{U_b}}{1+\frac{r\cdot\tilde{\ell}}{U_b}} \cdot \frac{1}{s_b} \cdot \log\left(\frac{s_b}{U_b + r\cdot\tilde{\ell}}\right) \, , \tag{242}$$

where we have modified the establishment time to interpolate over the transition to recombination dominated dynamics.

Once recombination represents the dominant mode of fixation, the sojourn time to reach a majority can become even larger, scaling with the period of oscillations T_{per}. This is because a mutation's chance at fixing will be capped at the period over which the transfer rate declines. Once a new class of replenishing mutations attains a majority of the nose, it will rapidly push minority frequency older mutations extinct. Since mutations continue to jackpot roughly uniformly over the duration of a period, we can simply estimate the sojourn time to attain a majority frequency as,

$$\langle \tau_{1\to N/2} \rangle \approx \frac{1}{\gamma^*} \sum_{m=0}^{\gamma^*} \sum_{n=1}^{q(0)} (q(0)+n+m) \cdot \left(1-\frac{1}{q(0)}\right)^{n-1} \cdot \left(\frac{1}{q(0)}\right) \cdot \tau_q(n) \sim (2q(0)+\gamma^*) \cdot \langle \tau_{qr} \rangle \, . \tag{243}$$

This time can be substantially larger than the asexual sojourn time, highlighting how sex can increase the sojourn times of mutations. Since we have assumed $\log(r\cdot\tilde{\ell}/U_b) \ll \log(Ns_b)$ to approximate $\gamma^* \ll 2\log(Ns_b)$, this sojourn time is capped as,

$$\langle \tau_{1\to N/2} \rangle \lesssim \frac{1}{s_b} \cdot \log(Ns_b) \cdot \log^{1/2}(s_b/r\cdot\tilde{\ell}) \, , \tag{244}$$

which represents an asymptotically larger value than that in the rapid recombination regime where $\langle \tau_{1\to N/2} \rangle \sim 1/s_b \log(Ns_b)$.

On the other hand, majority frequency mutations in sexual populations will fix faster than those in asexual population. This is because once a sexual mutation attains a majority of the population, it will grow deterministically like that in an asexual population. As we have shown, this occurs because large frequency mutations in sexual populations become a dominant acceptor background and grow by deterministically drafting to larger frequencies. We can ask how much a large frequency mutation grows over a single transfer cycle by converting the growth kernel of an acceptor mutation after one transfer $\rho(h|h_0)$ in Eq. (199) into logit-coordinates,

$$\rho(z|z_0) \approx \frac{\sin(\pi\alpha)/2\pi}{\cosh(\alpha_1 z - z_0/\alpha^{q_k-1}) + \cos(\pi\alpha)} \, . \tag{245}$$

This distribution is symmetric around $z \approx z_0/\alpha^q$, which corresponds to the typical growth factor of an asexual mutation over one nose-to-mean timescale and q establishments. This shows that majority frequency sexual and asexual mutations will typically increase by the same relative amount over the same number of establishments. However, sexual mutations will have a slightly shorter establishment time, so their overall sojourn time to fix after jackpotting will be relatively smaller by a factor proportional to $r \cdot \tilde{\ell}/U_b$,

$$\tau_{N/2 \to N} \approx \frac{1}{\gamma^*} \sum_{\gamma=0}^{\gamma^*} \frac{.577 + \log(\log(N))}{|\log(\alpha_\gamma)|} \cdot \frac{1}{s_b} \cdot \log\left(\frac{s_b}{r \cdot \tilde{\ell}}\right) \sim \frac{.577 + \log(\log(N))}{s_b} \log\left(\frac{s_b}{r \cdot \tilde{\ell}}\right) . \quad (246)$$

10.14 Regime of validity

These solutions and the single transfer regime will remain valid provided $v \ll s_b^2$,

$$2\log(N s_b) \ll \log^2(s_b/r \cdot \tilde{\ell}) . \quad (247)$$

When this condition breaks down, several aspects of evolutionary dynamics will change from the picture we have described above. First, the bulk will broaden to include fitness classes above the mean fitness. As a result, mutations will begin to transfer to higher fitness backgrounds before they reach the mean fitness. Moreover, the "nose" class of successful acceptor backgrounds will expand beyond the fitness class directly below x_c. This will coincide with a breakdown of the clonal dynamics we have modeled here and represent a transition to the diffusive dynamics characteristic of the rapid recombination regime. Future work to understand this transient regime where $\log^{1/2}(N s_b) \gg \log(s_b/r \cdot \tilde{\ell}) \gg 1$ will likely involve extending the 'moderate speeds regime' explored in prior asexual theory (57) (56).

11 Recombination dominated dynamics: colocalized transfer regime

We are now in a position to extend these results to the regime where asexuality breaks down due to multiple colocalized pairs attaining a large frequency. As we study these dynamics, we expect they will partially resemble those of the single transfer regime. This is because the total transfer rate of colocalized pairs will also fluctuate over time, and consequently, we will need to understand how this influences the rate of adaptation and width of the fitness distribution over time. There will also be several extensions beyond the single transfer regime. Colocalized pairs will jump multiple fitness classes when they establish in any nose. As a result, we will need to extend our analysis beyond the feeding class directly below the nose. As we will show, the competition between new mutations and transferred pairs can both lead to oscillations in the size of the nose over extremely short timescales and allow the transfer rate to reach extremely small values. To begin to consider these dynamics, we outline the competition between new and transferred mutations in the nose.

11.1 Branching process dynamics in the nose

The new nose will now be predominantly composed of three subclasses. The first subclass will consist of new mutations f_m produced by the feeding class directly below the nose $x = x_c - s_b$. Secondly, there will also be a subclass of transferred single mutations f_{r1} with an acceptor directly below the nose $x = x_c - s_b$. And lastly, there will be a subclass of transferred colocalized pairs f_{r2} with an

acceptor two fitness classes below the nose $x = x_c - 2s_b$. The first two subclasses will obey the same branching processes in Eq. (141) and Eq. (175) that we have previously discussed,

$$\frac{df_{r1}(x_c,t)}{dt} = x_c(t)f_{r1}(x_c) + \frac{r\cdot\tilde{\ell}}{2}\cdot\underbrace{\sum_i [h(g_i|0,t)\cdot(1-h(g_i|x_c-s_b,t)]\cdot f(x_c-s_b,t)}_{\omega(t)} + \sqrt{\frac{f_{r1}(x_c)}{N}}\eta(t)\ , \tag{248}$$

$$\frac{df_m(x_c,t)}{dt} = x_c(t)f_m(x_c) + U_b\cdot f(x_c-s_b,t) + \sqrt{\frac{f_m(x_c)}{N}}\eta(t)\ , \tag{249}$$

where $\omega(t)$ continues to denote the transfer rate of single mutations. Provided $x_c(t) - 2s_b \gg \sqrt{v(t)}$, the subclass of transferred pairs f_{r2} will obey a new linear branching process that depends on the overall frequency of colocalized pairs $h(g_i, g_j)$ in the population,

$$\frac{df_{r2}(x_c,t)}{dt} = x_c(t)f_{r2}(x_c) + \frac{r\cdot\tilde{\ell}}{2}\cdot\underbrace{\sum_{i,j} [h(g_i,g_j|0,t)\cdot(1-h(g_i|x_c-2s_b,t)-h(g_j|x_c-2s_b,t)]\cdot f(x_c-2s_b,t)}_{\omega_2(t)} + \sqrt{\frac{f_{r2}(x_c}{N}} \tag{250}$$

where we have defined the transfer rate of colocalized pairs ω_2 and again assumed $s_b \gg \sqrt{v}$ so that the typical donor is at the bulk of the population. The acceptor class for single transfers will again be exponentially growing,

$$f(x_c - s_b)(t) = \frac{e^{\int_0^t dt'(x_c(t)-s_b)}}{2Nx_c(t)}\ . \tag{251}$$

The acceptor class for colocalized transfers will grow at a slightly slower rate; however, it will begin at a larger size proportional to the amount it grew one establishment earlier,

$$f(x_c - 2s_b)(t) = \frac{e^{(x_c-s_b)\tau_q' + \int_0^t dt'(x_c(t)-2s_b)}}{2Nx_c(t)}\ , \tag{252}$$

which depends on the establishment time of this earlier class τ_q'. These branching processes highlight several important dynamics. First, like in the single transfer regime, the transfer rate of colocalized pairs will depend on the underlying genetic diversity. As the frequencies of colocalized pairs in the population varies, so too can the establishment time in the nose. To understand this process, we will again need to follow the nose over multiple nose-to-mean timescales. Secondly, there is now a feedback in the nose over the much shorter timescale of individual establishments. This occurs because the establishment time of the subclass founded by colocalized pairs $\tau_{q,r2}$ depends on the previous establishment time of the whole nose τ_q'.

$$\tau_{q,r2} \approx \frac{1}{(q-2)s_b}\log\left(\frac{s_b}{r\cdot\tilde{\ell}\cdot\omega_2(t)}\right) - \frac{(q-1)}{(q-2)}\cdot\tau_q'\ . \tag{253}$$

As a result, an atypically early or late nose establishment will immediately exert an influence on the next establishment time. To understand this feedback and how genetic diversity varies over time, we will again trace the dynamics forward in time, beginning from an asexual starting condition. This allows us to suddenly turn recombination on and ask whether new steady state behavior emerges. To further simplify the dynamics, we will focus on the regime where $r\cdot\tilde{\ell} \ll U_b$. This allows us to neglect the

contributions of single transfers and focus on the competition between the subclass founded by colocalized transfers and new mutations. Continuing with the convention introduced in the single transfer regime, we will continue to index time by nose to mean timescales since the asexual initial condition with γ.

11.2 Initial sexual dynamics

Provided, the population has been evolving asexually, we can make several key starting assumptions. First, the overall frequency of colocalized pairs in the bulk will simply be proportional to their smaller feeding rate $\sim U_b \cdot \tilde{\ell}/L$. This allows us to approximate the typical initial frequency of colocalized pairs in the bulk donor class,

$$f_p(\gamma = 0) \equiv \frac{f_{r2}(x = 0, \gamma = 0)}{f_m(x = 0, \gamma = 0)} \approx \left(\frac{\tilde{\ell}}{L} \right)^{1+\frac{1}{q_a}} . \tag{254}$$

Secondly, we can also assume the most recent establishment time τ_q' will coincide with the asexual establishment time $\tau_{q,m}$ defined in Eq. (141). Substituting this value into Eq. (253) allows us to straightforwardly estimate the establishment time of the recombinant subclass,

$$\tau_{q,r2}(\gamma = 0) \sim \frac{1}{(q_a - 2)s_b} \log \left(\frac{U_b}{r \cdot \tilde{\ell} \cdot f_p(0)} \right) , \tag{255}$$

and the relative fraction of the next nose founded by transferred pairs versus mutations,

$$f_p(1) \equiv \left\langle \frac{f_{r2}(x_c, 0)}{f_m(x_c, 0)} \right\rangle \sim e^{x_c(\tau_{q,r2}(0) - \tau_{q,m})} \approx \left(\frac{s_b \cdot r \cdot \tilde{\ell} \cdot f_p(0) \cdot \omega_2(0)}{U_b^2} \right)^{1+\frac{2}{q_a}} \cdot \left(\frac{U_b}{s_b} \right)^{\frac{1}{q_a}} \sim \left(\frac{s_b \cdot r \cdot \tilde{\ell} \cdot f_p(0)}{U_b^2} \right) , \tag{256}$$

where $\omega_2(0) \sim \mathcal{O}(1)$ since the original class of colocalized pairs has not experienced genetic drowning. To begin to get a sense for how this first transfer event affects the frequency of colocalized pairs, we can substitute the critical value of the transfer rate where colocalized pairs start to dominate from Eq. (101) into Eq. (256). At this crossover rate, the relative fraction in Eq. (256) will be equivalent to the value in Eq. (254). This demonstrates that as we move to the recombination dominated regime, colocalized pairs in the nose will deterministically increase in frequency over multiple transfer cycles. It also allows us to straightforwardly interpret the breakdown of quasi-asexuality as the critical transfer rate where colocalized pairs deterministically grow in the nose.

As the initial class of colocalized pairs drives adaptation, it will experience genetic drowning analogous to that in the single transfer regime. Provided the total number of colocalized loci in the population is large $\tilde{\ell} \cdot f_b \gg 1$, we can again show that the transfer rate will decline systematically with each transfer,

$$\omega_2(\gamma) \sim \prod_{n=0}^{\gamma} \left(1 - \frac{2}{q(n)} \right) , \tag{257}$$

where $q(0) \equiv q_a$ now corresponds to the asexual value. We can now ask how many transfers until the transfer rate reaches its maximum value given the asexual starting condition.

11.3 Time for colocalized pairs to reach large frequency

By setting Eq. (256) equal to one, we can see that once the relative fraction of colocalized pairs attains the small value,

$$f_p(\gamma_1^*) \approx \left(\frac{U_b^2}{r\cdot\tilde{\ell}\cdot\omega_2(\gamma_1^*)\cdot s_b}\right)\cdot\left(\frac{s_b}{U_b}\right)^{1/q_a} , \tag{258}$$

colocalized pairs will reach $\mathcal{O}(1)$ in the next transfer cycle. As colocalized pairs vie to reach this critical frequency, they will compete against a predominantly asexual population. This allows us to estimate the number of transfer cycles γ_1^* until this frequency is reached using the simple update rule in Eq. (256),

$$0 \approx \left(1+\frac{2}{q_a}\right)^{\gamma_1^*}\cdot\log(f_p(0)) - \frac{2}{q_a}\left(1+\frac{2}{q_a}\right)\sum_{n=0}^{\gamma_1^*-1}(\gamma_1^*-1-n)\cdot\left(1+\frac{2}{q_a}\right)^n \tag{259}$$

$$+\left[\left(1+\frac{2}{q_a}\right)\cdot\log\left(\frac{U_b^2}{r\cdot\tilde{\ell}\cdot s_b}\right)+\frac{1}{q_a}\log\left(\frac{U_b}{s_b}\right)\right]\cdot\sum_{n=0}^{\gamma_1^*-1}\left(1+\frac{2}{q_a}\right)^n . \tag{260}$$

We can straightforwardly evaluate these sums and solve for γ_1^* to show that colocalized pairs will attain $\mathcal{O}(1)$ frequency in the nose after,

$$\gamma_1^* \approx \max\left(\frac{q_a}{2}\cdot\log\left[\frac{\frac{q_a}{2}\log(\frac{r\cdot\tilde{\ell}\cdot s_b}{U_b^2})-\frac{1}{2}\log\left(\frac{s_b}{U_b}\right)}{\frac{q_a}{2}\log(\frac{r\cdot\tilde{\ell}\cdot s_b}{U_b^2})-\frac{1}{2}\log\left(\frac{s_b}{U_b}\right)+\left(1+\frac{1}{q_a}\right)\cdot\log\left(\frac{\tilde{\ell}}{L}\right)}\right],1\right) , \tag{261}$$

where we have approximated to first order in $q_a \gg 1$. Provided colocalized pairs dominate, we will have $q_a/2\log(r\cdot\tilde{\ell}\cdot s_b/U_b^2) \gg \log(L/\tilde{\ell})$, which allows us to expand Eq. (261) to show,

$$\gamma_1^* \approx \max\left(-\frac{\log(\tilde{\ell}/L)}{\log\left(\frac{r\cdot\tilde{\ell}\cdot s_b}{U_b^2}\right)},1\right) . \tag{262}$$

11.4 Oscillations in the nose over short timescales

Once colocalized pairs attain a $\mathcal{O}(1)$ frequency, they will drive the advance of the fitness distribution. As this occurs, the size of the nose will begin to dramatically oscillate over the short timescale of individual establishments. We can see this by continuing to iterate the dynamics forward. In particular, after $f_p(\gamma_1^*) \sim \mathcal{O}(1)$, the establishment time will be,

$$\tau_{q,r2}(\gamma_1^*) \sim \frac{1}{(q-2)s_b}\log\left(\frac{U_b}{r\cdot\tilde{\ell}\cdot\omega_2(\gamma_1^*)}\right) . \tag{263}$$

This constitutes a rapid establishment time compared to the asexual establishment time. As a result, the next colocalized acceptor class below the nose will have grown for grown for commensurately less time and be proportionally smaller. In particular, we can show that the next establishment time of the colocalized class in the nose will be,

$$\tau_{q,r2}(\gamma_1^*+1) \approx \frac{1}{(q-1)s_b}\log\left(\left(\frac{s_b}{U_b}\right)\cdot\left(\frac{r\cdot\tilde{\ell}\cdot\omega_2(\gamma_1^*)\cdot s_b}{U_b^2}\right)^{1/q}\right) . \tag{264}$$

This represents a substantially larger time than that in Eq. (263) and crucially a longer time than that of the mutation founded class. Consequently, the next nose will typically be largely composed of new mutations. In turn, this guarantees that the next establishment time will again be that in Eq. (263). This shows that nose will typically oscillate between extremely rapid establishments of colocalized pairs and slow establishments of new mutations,

$$\tau^{ss}_{q,m}(\gamma > \gamma^*_1) \approx \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b}\right) , \tag{265}$$

$$\tau^{ss}_{q,r2}(\gamma > \gamma^*_1) \approx \frac{1}{(q-2)s_b} \cdot \left(\frac{U_b}{r \cdot \tilde{\ell} \cdot \omega_2(\gamma)}\right) . \tag{266}$$

These oscillations are interesting for several reasons. First, they suggest that the shape of the fitness distribution can deviate from its Gaussian shape, developing heavy tails. This behavior has been predicted in previous work (7) and highlights another mechanism that causes the steady state traveling wave picture to fail. It also means that, unlike the single transfer regime, genetic variants in the colocalized pair dominated regime, can attain large frequencies on their original genetic background. Interestingly, these new mutations will typically have been leapfrogged by colocalized pairs in the class directly below them. This implies that despite attaining large overall frequencies, these new mutations will not typically leave future descendants, including introducing new colocalized pairs. This represents a surprising decoupling between which genetic are successful and which genetic variants are at a large frequency.

11.5 Unstable non-equilibrium oscillations in the nose

While the oscillations between the establishment times in Eq. (265) and Eq. (266) represent the typical dynamics, random early establishments can disrupt this oscillatory behavior. We briefly show that these disruptions are unstable and rapidly return to the typical dynamics outlined above. After an atypically early or late establishment τ'_q, the typical next establishment time can be obtained from Eq. (253),

$$\tau_{q,r2} \approx \tau^* + \left(1 + \frac{1}{q}\right)\left[\tau^* - \tau'_q\right] , \tag{267}$$

where

$$\tau^* \equiv \frac{1}{(2q-3)s_b} \log\left(\frac{s_b}{r \cdot \tilde{\ell} \cdot \omega_2}\right) . \tag{268}$$

This time represents an intermediate value between Eq. (265) and Eq. (266), ensuring that colocalized pairs will dominate every establishment. As this occurs, we can extend the recursive dependence Eq. (267) to calculate the typical establishment time n establishments after the atypical early establishment τ'_q,

$$\tau_{q,r} \approx \tau^* + (-1)^n \cdot \left(\frac{q-1}{q-2}\right)^{n+1} \left[\tau^* - \tau'_q\right] . \tag{269}$$

This relationship highlights that the establishment time will unstably oscillate away from τ^* until the stable steady state values are attained. Even for extremely atypical establishments, where $|\tau^* - \tau'_q| \sim 1/qs_b$, we can see that that nose will attain stable steady oscillations within $\mathcal{O}(q(0))$ establishments by setting Eq. (269) equal to Eq. (265). This implies that unstable oscillations typically revert within a single transfer cycle.

11.6 Width of the fitness distribution

In reality, establishment times are distributed around a typical value and the recurrence relation in Eq. (269) reveals that any early or late establishment in the past compared to τ^* will drive increasingly large oscillations away from τ^*. Eventually, these oscillations will get so large that $\tilde{\tau}_{q,r} \approx \tilde{\tau}_{q,m}$ and mutation founded clones will start to dominate the new nose class. We can now ask how the dynamics change once this happens.

To understand how the fluctuations in the nose influence the width of the fitness distribution, we integrate over the growth of a new nose class until it reaches the bulk. Since the transfer rate will typically only decline by a factor $e^{-1} \sim \mathcal{O}(1)$ during the nose to mean timescale, we can approximate it as roughly constant over this span,

$$N \approx \frac{e^{\sum_{n=1}^{q(\gamma)/2-1}(\frac{q(\gamma)}{2}-n)\cdot 2s_b\cdot(\tau^{ss}_{q,m}+\tau^{ss}_{q,r2})}}{q(\gamma)\cdot s_b} \,, \tag{270}$$

Substituting our values for the establishment times and solving for $q(\gamma)$, we obtain,

$$q(\gamma) \sim \frac{4\log(Ns_b)+\log(s_b/U_b)}{\log(s_b/r\cdot\tilde{\ell}\cdot\omega_2(\gamma))} \sim q(\gamma) \sim \frac{4\log(Ns_b)-k}{\log(s_b/r\cdot\tilde{\ell})} \,. \tag{271}$$

This demonstrates, that in spite of the large oscillations occurring in the nose, we can still continue to roughly approximate the width of the fitness distribution as equivalent to the asexual value with an effective fitness benefit $s_b \to 2s_b$ and a time-dependent effective mutation rate $U_b \to r\cdot\tilde{\ell}\cdot\omega_2(\gamma)$.

11.7 Exhausting colocalized pairs over long timescales

As the population continues to adapt, genetic drowning will drive the transfer rate of colocalized pairs to lower values. As a result, the value of $q(\gamma)$ will commensurately decrease with each transfer γ since the initial condition. This allows us to extend our approximations from the single transfer regime to show,

$$q(\gamma) \approx q(0)\cdot e^{-\frac{\gamma}{2\log(Ns_b)}} \,, \tag{272}$$

and,

$$\omega(\gamma) \approx e^{-\frac{2\gamma}{q(0)}} \,. \tag{273}$$

Now, to predict when the initial cohort of colocalized pairs is exhausted we will have to consider two possibilities. First, like in the single transfer regime, small subclasses of new colocalized pairs could accumulate until they're suddenly able to dominate the transfer rate. However, we can already see that these subclasses will be dramatically smaller when first produced compared to the single transfer regime,

$$\left\langle \frac{f_{r2}(x_c,\gamma)}{f_{r2}(x_c,0)} \right\rangle \sim \frac{r\cdot\tilde{\ell}\cdot\omega(\gamma)\cdot s_b}{(U_b\cdot\tilde{\ell}/L)^2} \,. \tag{274}$$

By generalizing our arguments from the single transfer regime, we can see that this this leads to a predicted number of transfers in a period γ_2^*,

$$\gamma_2^* \sim \sqrt{2q(0)\cdot\log\left(\frac{r\cdot\tilde{\ell}\cdot s_b}{(U_b\cdot\tilde{\ell}/L)^2}\right)} \,, \tag{275}$$

provided $4q(0) \gtrsim \log\left(\frac{r\cdot\tilde{\ell}\cdot s_b}{(U_b\cdot\tilde{\ell}/L)^2}\right)$. Alternatively, the nose could simply become dominated by new mutations after a characteristic time,

$$\gamma_2^* \sim \frac{q(0)}{2}\cdot\log\left(\frac{r\cdot\tilde{\ell}\cdot s_b}{U_b^2}\right) - \frac{1}{2}\log\left(\frac{s_b}{U_b}\right) - \gamma_1^* \,. \tag{276}$$

In fact, it is straightforward to see by substituting in the critical value of $r\cdot\tilde{\ell}$ from Eq. (101) into Eq. (275) and Eq. (276) that Eq. (276) will initially represent a much smaller timescale and thus the typical length of a period before new mutations start to dominate in the nose, resetting our asexual initial condition. This indicates a clear difference between the single a colocalized pair transfer regimes. In the single transfer regime new mutations do not attain detectable frequencies on their original backgrounds. By contrast, when colocalized pairs dominate, populations can undergo spans of entirely asexual evolution after diversity crashes.

Long term estimates of the rate of adaptation Using these results, we can estimate the rate of adaption by summing over the amount the population increases in fitness over each transfer cycle and the corresponding time,

$$v \approx \frac{\sum_{\gamma=0}^{\gamma_1^*} q_a\cdot s_b + \sum_{\gamma_1^*}^{\gamma_1^*+\gamma_2^*} q(\gamma)\cdot s_b}{\sum_{\gamma=0}^{\gamma_1^*} q_a\cdot\tau_{q,m} + \sum_{\gamma_1^*}^{\gamma_1^*+\gamma_2^*}\frac{q(\gamma)}{2}\cdot(\tau_{q,r}^{ss}(\gamma)+\tau_{q,m}^{ss})} \tag{277}$$

Substituting for these establishment times we obtain,

$$v \approx s_b^2\cdot\frac{\gamma_1^*\cdot q_a + \gamma_2^*\cdot q(0) - \frac{(\gamma_2^*+1)(2\gamma_1^*+\gamma_2^*)}{\log(s_b/r\cdot\tilde{\ell})}}{\gamma_1^*\cdot\log(s_b/U_b) + \frac{\gamma_2^*}{2}\cdot\log(s_b/r\cdot\tilde{\ell}) + \frac{(\gamma_2^*+1)(2\gamma_1^*+\gamma_2^*)}{2q(0)}} \,, \tag{278}$$

To first order in $4\log(s_b/r\cdot\tilde{\ell}) \gg \log(r\cdot\tilde{\ell}\cdot s_b/U_b^2)$, this simplifies to,

$$v \approx \frac{q^2(0)s_b^2}{2\log(Ns_b)}\cdot\left[\frac{\frac{\gamma_1^*}{\gamma_2^*}\cdot\frac{q_a}{q(0)} + \left(1-\frac{2\gamma_1^*+\gamma_2^*}{q_0\cdot\log(s_b/r\cdot\tilde{\ell})}\right)}{2\frac{\gamma_1^*}{\gamma_2^*}\cdot\frac{q(0)}{q_a} + \left(1+\frac{2\gamma_1^*+\gamma_2^*}{q_0\cdot\log(s_b/r\cdot\tilde{\ell})}\right)}\right] \,. \tag{279}$$

This suggests two asymptotic approximations for the rate of adaptation. when $\gamma_1^* \gg \gamma_2^*$ or $-2/q_a\log(\tilde{\ell}/L) \gg \log^2(s_b\cdot r\cdot\tilde{\ell}/U_b^2)$, we can solve perturbatively around the asexual rate of adaptation,

$$v \approx v_a\left[1+\frac{\log\left(\frac{r\cdot\tilde{\ell}\cdot s_b}{U_b^2}\right)}{-\frac{2}{q_a}\log\left(\frac{\tilde{\ell}}{L}\right)}\right] \,. \tag{280}$$

In the opposite regime, we can expand around $\gamma_2^* \gg \gamma_1^*$ to obtain,

$$v \approx \frac{2\cdot q^2(0)s_b^2}{2\log(Ns_b)}\cdot\left[1-\frac{\log\left(\frac{r\cdot\ell\cdot\tilde{s_b}}{U_b^2}\right)}{2\log\left(\frac{s_b}{r\cdot\tilde{\ell}}\right)}\right] \,. \tag{281}$$

The results also allow us to simply approximate the relative fraction of mutations fixing on their original background as,

$$\log\left(\frac{\theta_r}{\theta_a}\right) - 1 \sim \frac{1}{1 + \frac{\log\left(\frac{r\cdot\ell\cdot\tilde{s}_b}{U_b^2}\right)}{-\frac{2}{q_a}\log\left(L/\tilde{\ell}\right)}} \,. \tag{282}$$

12 Distributions of transfer lengths

Thus far, we have focused on a simplified model where all transferred segments have the same genetic length. However, most organisms can incorporate segments than span a distribution of genetic lengths $\lambda(\ell)$. It is intriguing to ask whether certain segments are more likely to contribute to the success of recombined variants. The key challenge is that we can no longer consider separate colocalized and non-colocalized branching processes. This is because the probability that a mutation is transferred with its colocalized neighbor will depend on the tail of $\lambda(\ell)$,

$$\tilde{R}_{2\to} = r \cdot \int_{\delta\ell_2}^{L} \lambda(\ell)\cdot(\ell - \delta\ell_2)\, d\ell \,. \tag{283}$$

However, provided the tail of $\lambda(\ell)$ falls off at least exponentially fast, we will be able to approximate.

$$\tilde{R}_{2\to} \approx \begin{cases} r\cdot\overline{\ell} & \text{if } \delta\ell_2 \lesssim \mathcal{O}(\overline{\ell}) \,, \\ 0 & \text{if } \delta\ell_2 \gtrsim \mathcal{O}(\overline{\ell}) \,. \end{cases} \tag{284}$$

This allows us to continue to approximate the boundaries of the regime where single transfers and colocalized pairs dominate by substituting $\tilde{\ell} \to \overline{\ell}$. We consider these scenarios separately below

Single transfers Provided quasi-asexuality breaks down due to the transfer of individual mutations, we have shown that the substitution rate of recombined mutations will scale linearly with the transfer rate $\theta_r \propto r\int\lambda(\ell)\cdot\ell$. This is because recombination events will primarily involve individual genetic variants and so the transfer rate will scale with the ℓ possible start sites. As a result, we find that the distribution of successful transfer lengths scales linearly with ℓ,

$$\lambda(\ell)_{\text{fixed}} \propto \lambda(\ell)\cdot\ell \,. \tag{285}$$

Colocalized transfers In larger populations $(NU_b)^{3/4} \gtrsim L/\overline{\ell}\cdot\tilde{k}$, the dominant contribution will come from the repeated transfer of colocalized mutations. A new successful pair will arise with a spacing drawn from $\Pr(\delta\ell_2)$. This pair can be transferred by any segment with length $\ell > \delta\ell_2$. As a result, the distribution of successful genetic lengths will scale as,

$$\lambda(\ell)_{\text{fixed}} \equiv \int_0^{\ell} \underbrace{\left[\frac{\lambda(\ell)\cdot(\ell-\delta\ell_2)}{\int_{\delta\ell_2}^{L}\lambda(\ell')\cdot(\ell'-\delta\ell_2)d\ell'}\right]}_{\Pr(\ell \text{ contributed to fixation of pair})} \cdot \underbrace{\left[\frac{Pr(\delta\ell_2)\cdot\ \theta_r(\delta\ell_2)}{\int_0^{L}\theta_r(\delta\ell_2')d\delta\ell_2'}\right]}_{\Pr(\text{spacing btw. fixed pair})} d\delta\ell_2 \,, \tag{286}$$

where as we have shown,

$$\theta_r(\delta\ell_2) \propto (\tilde{R}_{2\to})^{\tilde{k}/4} \,. \tag{287}$$

Provided the tail of $\lambda(\ell)$ falls of at least exponentially fast, we will be able to approximate,

$$\lambda(\ell)_{\text{fixed}} \propto \lambda(\ell) \cdot \int_0^{\ell} (\ell - \delta\ell_2)^{1+\frac{\tilde{k}}{4}} \, \frac{d\delta\ell_2}{\ell} \, . \tag{288}$$

This yields a distribution of successful transfer lengths that to leading order scales quadratically with ℓ,

$$\lambda(\ell)_{\text{fixed}} \propto \lambda(\ell) \cdot \ell^2 \, . \tag{289}$$

Single-ℓ effective model For many shapes of $\lambda(\ell)$ (e.g. an exponential distribution) the fixed distributions will be centered around a characteristic length. As a result, we find that we can understand these scenarios with our single-ℓ theory by defining $\tilde{\ell}$ to coincide with this value. For many simple shapes of $\lambda(\ell)$, $\tilde{\ell}$ will occur near the mean length of a transferred segment $\bar{\ell}$. In this scenario the genetic lengths of successful segments will not vary with the population size. However, HGT can often involve multiple biological sources and length scales, in which case $\lambda(\ell)$ may have multiple peaks. In these scenarios, we find that the average length of a transferred segment can be a poor predictor for what is successful.

13 Distributions of Fitness Effects

While we have focused on a simple model of the fitness landscape where all beneficial mutations confer the same fitness benefit, organisms can produce mutations with a range of fitness benefits. These scenarios will complicate our picture since we must jointly keep track of the fitness benefits of mutations and their frequency. To begin to understand these scenarios, we can extend our model to allow the fitness effects of mutations along the genome to be randomly sampled from a continuous distribution of fitness effects, $\mu(s)$,

$$\mu(s) \equiv U_b \cdot \rho(s) \, , \tag{290}$$

where we continue to assume that each locus mutates at the same rate U_b/f_bL. The resulting dynamics are simplest to understand in the quasi-asexual regime, where we can leverage previous work that describes the dynamics of asexual populations with a continuous DFE. In particular, this work has shown that the distribution of fixed mutations $\rho_{fix}(s)$ will be determined by the values of s that dominate the integral

$$\rho_{fix}(s) \propto \int_0^{\infty} \rho(s) \cdot e^{\frac{x_{ca}s}{v_a} - \frac{s^2}{2v_a}} \left[\frac{1}{s} + \frac{1}{x_{ca} - s} \right] \, ds \, , \tag{291}$$

Even when the DFE is widely distributed, this integral will often be concentrated around a characteristic fitness benefit $\tilde{s_b}$, which occur at an overall rate $\tilde{U_b}$. This yields a general condition for x_c,

$$1 = \frac{\tilde{U_b}}{\tilde{s_b}} \left[1 - \frac{\tilde{s_b}}{x_{ca}} \right]^{-1} e^{\frac{x_{ca}\tilde{s_b}}{v_a} - \frac{\tilde{s_b}^2}{2v_a}} \, , \tag{292}$$

where s_b satisfies the condition

$$x_c - \tilde{s_b} + v_a \frac{\partial \log \mu(\tilde{s_b})}{\partial s} = 0 \, . \tag{293}$$

The overall mutation rate of these mutations will be,

$$\tilde{U_b} = \sqrt{2\pi\Delta s_b}\mu(\tilde{s_b}) \, , \tag{294}$$

where,

$$\Delta s_b = \left[\frac{1}{v} - \frac{\partial^2 \log \mu(\tilde{s_b})}{\partial s^2}\right]^{-1/2} . \tag{295}$$

These approximations allow us to estimate the substitution rate of linked mutation,

$$\theta_a \approx N\tilde{U_b} \cdot e^{\frac{(x_c - \tilde{s_b})^2}{2v}} . \tag{296}$$

These results also crucially allow us to assume that most mutations at a large frequency also have the same fitness benefit $\tilde{s_b}$. We can therefore straightforwardly extend our branching process for mutations that are not initially colocalized $w_{r>}(x)$ provided we substitute $\tilde{s_b} \to s_b$ and $\tilde{U_b} \to U_b$. This allows us to calculate the substitution rate of single mutations and colocalized pairs.

Contributions from single transferred mutations In small populations the dominant contribution to the substitution rate will arise from recombined mutations that transfer a single time. This allows us to simply estimate the fraction of mutations taking over on a different genetic background,

$$\frac{\theta_r}{\theta_a} - 1 \approx \frac{\tilde{k} \cdot r \cdot \tilde{\ell}}{\tilde{U_b}} . \tag{297}$$

Since this effective mutation rate is dominated by mutations in the tail it can be more orders of magnitude smaller than the actual overall beneficial mutation rate. This suggests that continuous distributions of fitness effects can drive recombination to play a role in evolutionary dynamics at far smaller rates. In larger populations, the typical mode of fixation will involve multiple colocalized mutations. The important extension is that we now must generalize the contributions from producing a colocalized mutation to consider a continuous value of s.

Contributions from colocalized pairs Rarely, a new mutation will arise near another preexisting mutation, which we assume has a fitness benefit of $\tilde{s_b}$. This pair can then be repeatedly transferred as it competes against a largely asexual population to take over. The crucial extension is that the fixation probability of this pair will depend on the total fitness benefit of the pair: $\tilde{s_b} + s$, where s is sampled from $\mu(s)$. This implies that we now must extend $w_{r<}(x, \delta\ell_2)$ to be a function of s,

$$w_{r<}(x, \delta\ell_2) \to w_{r<}(x, \delta\ell_2, s) . \tag{298}$$

This allows us to extend the substitution rate of colocalized pairs

$$\theta_r(s) \approx N \int \mu(s)\, ds \int_0^{\tilde{\ell}} \Pr(\delta\ell_2)\, d\delta\ell_2 \int f(x) \cdot w_{r<}(x+s, \delta\ell_2, s)\, dx , \tag{299}$$

where $w_{r<}(x, \delta\ell_2, s)$ will satisfy the shoulder solution with an interference threshold,

$$1 \approx \frac{\tilde{R'}}{n\tilde{s_b}} \left(1 - \frac{n\tilde{s_b}}{x_{cr<}}\right)^{-1} \cdot e^{\frac{x_{cr<} n\tilde{s_b}}{v} - \frac{(n\tilde{s_b})^2}{2v}} , \tag{300}$$

where we have defined $n = (s + \tilde{s_b})/\tilde{s_b}$. An iterative solution yields,

$$x_{cr<} \approx \frac{1}{n}\left[x_c + \frac{\tilde{s_b}}{2}\left(n^2 - 1\right) + \frac{v}{\tilde{s_b}} \log\left(\frac{\tilde{U_b}}{\tilde{R'}}\right)\right] . \tag{301}$$

This allows us to evaluate the substitution rate,

$$\theta_r(n) \approx N \cdot \frac{\tilde{k} \cdot \tilde{\ell}}{L} \int_1^\infty dn\ \mu(\tilde{s_b} \cdot (n-1)) \cdot \tilde{s_b} \cdot e^{\left(1-\frac{1}{n^2}\right)\left(\frac{x_c^2}{2v} - \frac{x_c \tilde{s_b}}{2v}\right) + \frac{x_c}{\tilde{s_b}} \frac{1}{n^2} \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) - \frac{v}{2\tilde{s_b}^2 \cdot n^2} \log^2\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) + \frac{1}{2}\left(1-\frac{1}{n^2}\right) \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right)}\ , \tag{302}$$

This suggests n will be peaked a characteristic value determined by the condition,

$$\frac{\partial}{\partial n} \log[\mu(\tilde{s_b}(n^*-1)] + \frac{2}{n^{*3}} \cdot \left[\frac{x_c^2}{2v} - \frac{x_c \tilde{s_b}}{2v} - \frac{x_c}{\tilde{s_b}} \cdot \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) + \frac{v}{2\tilde{s_b}^2} \cdot \log^2\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) + \frac{1}{2} \cdot \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right)\right] = 0\,. \tag{303}$$

We can also estimate the fraction of mutations fixed on their original background by substituting Eq. (296),

$$\frac{\theta_r}{\theta_a} - 1 \approx \cdot \frac{\tilde{k} \cdot \tilde{\ell}}{L}\ \frac{\mu(\tilde{s_b} \cdot (n^*-1))}{\tilde{U_b}} \cdot \tilde{s_b} \cdot e^{\left(1-\frac{1}{n^{*2}}\right)\left(\frac{x_c^2}{2v} - \frac{x_c \tilde{s_b}}{2v}\right) + \frac{x_c}{\tilde{s_b}} \frac{1}{n^{*2}} \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) - \frac{v}{2\tilde{s_b}^2 \cdot n^{*2}} \log^2\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) + \frac{1}{2}\left(1-\frac{1}{n^{*2}}\right) \log\left(\frac{r \cdot \tilde{\ell}}{\tilde{U_b}}\right) - \frac{(x_c - \tilde{s_b})^2}{2v}}\ . \tag{304}$$

13.1 Specific distributions of fitness effects

Motivated by empirical observations and previous work, we consider a simple family of exponentially bounded distributions with a cutoff at $s = s_c$.

$$\mu(s) \equiv \begin{cases} \frac{U_b}{\sigma} \cdot \frac{e^{-(s/\sigma)^\beta}}{\Gamma(1+\beta^{-1})} \text{ if } s < s_c\ , \\ 0 \text{ else}\ . \end{cases} \tag{305}$$

In our simple additive model, this cutoff is motivated by the finite number of beneficial loci in the genome. Specifically, given $f_b \cdot L$ beneficial loci, the largest sampled value from Eq. (305) will typically be,

$$\text{argmax}[\mu(s)] \equiv s_c \approx \sigma \cdot \log^{1/\beta}(f_b \cdot L)\ . \tag{306}$$

This demonstrates that for $\beta \gg 1$, values larger than σ will be rarely sampled and thus play a negligible role in adaptation. In the opposite scenario where $\beta \sim 1$, this suggests there is a finite cutoff well above the mean.

For the distribution in Eq. (305), previous work has shown,

$$\tilde{s_b} \approx \begin{cases} \min(x_c - \frac{v}{\sigma}, s_c) \text{ if } \beta \sim 1\ , \\ \sigma/2 \text{ if } \beta \gg 1\ . \end{cases} \tag{307}$$

and

$$\tilde{U_b} \approx \begin{cases} U_b \cdot \sqrt{2\pi v} \cdot \frac{1}{\sigma} e^{-\frac{\tilde{s_b}}{\sigma}} \text{ if } \beta \sim 1\ , \\ U_b \text{ if } \beta \gg 1 \end{cases} \tag{308}$$

Previous work has also shown that an iterative solution for x_c yields,

$$x_c \sim \begin{cases} \sigma \cdot \log(NU_b) \text{ if } \beta \sim 1\ , \\ 2 \cdot \frac{\log(N\sigma)}{\log(\sigma/U_b)} \text{if } \beta \gg 1 \end{cases}\ . \tag{309}$$

This demonstrates the typical effect will be well-sampled for $\beta \gg 1$ provided $\log(NU_b) \lesssim \log(f_b \cdot L)$.

Substitution rate of single mutations With this family of distributions we can better understand the substitution rate of single mutations in Eq. (297). In particular when $\beta \sim 1$, this rate will scale as,

$$\log\left(\frac{\theta_r}{\theta_a}\right) \approx \log\left(\frac{r\cdot\tilde{\ell}\cdot\tilde{k}}{U_b}\right) + \min\{\log(NU_b), \log(f_b\cdot L)\}\ . \tag{310}$$

When $N\mu \ll 1$, the crossover to recombination will occur at extremely small recombination rates $r\cdot\tilde{\ell} \sim 1/N$. In the opposite limit where $N\mu \gg 1$, this crossover will saturate at $r/\mu \sim 1/\tilde{\ell}$. This demonstrates that the crossover to recombination dominated dynamics can depend on the population size even when single mutations are transferred. It also highlights that the crossover can occur well-below the threshold we might predict by naively substituting $\tilde{s_b} \approx \sigma$.

Substitution rate of colocalized pairs Substituting Eq. (305) into Eq. (303), we find that,

$$n^* \sim \begin{cases} \left[1+\frac{\sigma}{x_c}\cdot\log\left(\frac{\tilde{U_b}}{r\cdot\tilde{\ell}}\right)\right]^{1/3} \sim \left[\frac{\sigma}{x_c}\cdot\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)\right]^{1/3} \sim \left[\frac{\log\left(\frac{U_b}{r\cdot\tilde{\ell}}\right)}{\log(NU_b)}\right]^{1/3} \text{ if } \beta\sim 1\ , \\ 2-\mathcal{O}(1/\beta) \text{ if } \beta \gg 1\ . \end{cases} \quad . \tag{311}$$

This demonstrates that the typical fitness benefit of successfully recombined mutations will decrease as the recombination rate becomes larger. It also allows us to estimate,

$$\theta_r \approx N\cdot U_b\cdot\frac{\tilde{k}\cdot\tilde{\ell}}{L}\frac{\tilde{s_b}}{\sigma}\cdot\frac{e^{-\left(\frac{\tilde{s_b}\cdot(n^*-1)}{\sigma}\right)^\beta}}{\Gamma(1+\beta^{-1})}\cdot e^{\left(1-\frac{1}{n^{*2}}\right)\left(\frac{x_c^2}{2v}-\frac{x_c\tilde{s_b}}{2v}\right)+\frac{x_c}{\tilde{s_b}}\frac{1}{n^2}\log\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)-\frac{v}{2\tilde{s_b}^2\cdot n^{*2}}\log^2\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)+\frac{1}{2}\left(1-\frac{1}{n^{*2}}\right)\log\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)}\ . \tag{312}$$

We can also calculate the fraction of mutations taking over on a background different from their original one,

$$\frac{\theta_r}{\theta_a}-1 \approx \frac{\tilde{k}\cdot\tilde{\ell}}{L}\frac{\tilde{s_b}}{\sigma}\cdot e^{-\left(\frac{\tilde{s_b}\cdot(n^*-1)}{\sigma}\right)^\beta+\left(\frac{\tilde{s_b}}{\sigma}\right)^{1/\beta}}\cdot e^{\left(1-\frac{1}{n^{*2}}\right)\left(\frac{x_c^2}{2v}-\frac{x_c\tilde{s_b}}{2v}\right)+\frac{x_c}{\tilde{s_b}}\frac{1}{n^2}\log\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)-\frac{v}{2\tilde{s_b}^2\cdot n^{*2}}\log^2\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)+\frac{1}{2}\left(1-\frac{1}{n^{*2}}\right)\log\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)-\frac{(x_c-\tilde{s_b}}{2v}} \tag{313}$$

When $\beta \sim 1$, this simplifies to,

$$\log\left(\frac{\theta_r}{\theta_a}\right) \approx \log\left(\frac{\tilde{k}\cdot\tilde{\ell}}{L}\right) - \frac{2\tilde{s_b}\cdot(n^*-1)}{\sigma} + \frac{\tilde{s_b}(x_c-\tilde{s_b})}{2v} + \frac{1}{2}\cdot\log\left(\frac{r\cdot\tilde{\ell}}{\tilde{U_b}}\right)\ . \tag{314}$$

Since $n^* \geq 1$, the substitution rate of colocalized pairs will always be smaller than single mutations.

14 Incorporating deleterious mutations

While we have primarily focused on the effect of recombination on positive selection, the vast majority of new mutations will be deleterious. Previous work has shown that we can categorize deleterious mutations based on their selective effect. Mutations that are strongly deleterious $T_c|s_d| \gtrsim 1$, will seldom reach large frequency. More weakly selected mutations $T_c|s_d| \lesssim 1$ are quasi-neutral and can hitchhike to large frequencies with beneficial mutations. In principle recombination can provide a benefit to a clone by overwriting these mutations with the clonal background. When multiple of these mutations are colocalized, a genotype could obtain a multiplicative benefit.

In an asexual population, the total number of deleterious mutations at large frequency will be limited by their substitution rate,

$$\theta_d(s) \sim U_d \cdot e^{\frac{(x_c - s_b)s_d}{v}} . \tag{315}$$

Previous work has shown that for empirical estimates of the deleterious mutation rate and selection coefficient, these mutations will fix at a far lower rate than beneficial mutations. Since the fitness cost of these mutations is also lower than that of a beneficial mutation they will often play a negligible role in the breakdown of asexuality.

15 More general models of epistasis and genome organization

Thus far, we have focused on the simplest possible fitness landscape where 1) the number of beneficial sites remains constant over time 2) beneficial mutations are uniformly distributed along the genotype, and 3) the fitness benefits of mutations are simply additive. We are now in a position to relax these assumptions and understand how they influence our key results.

15.1 Diminishing returns epistasis

While we have assumed that the overall number of beneficial loci and their fitness benefits remain constant over time, a large body of empirical work evidences diminishing returns (44)(50). This process can occur through a large number of biological mechanisms. For example, multiple beneficial mutations may target the same cellular process and be redundant. Therefore, as mutations take over, the number of target sites may also decline. Alternatively, adaptation may simply be a coupon collecting process whereby beneficial mutations gradually run out over time. These processes would effect our model by making the beneficial fraction (and beneficial mutation rate) time dependent. In the simplest extension, the beneficial mutation rate would decline with the fixation rate $\theta(t)$ as mutations fix are not replaced.

$$\partial_t f_b(t) = \theta(t) \cdot f_b(t) . \tag{316}$$

Our results show that in the quasi-asexual regime only a moderate number of mutations tend to take over in a fixation timescale $\sim \mathcal{O}(\tilde{k})$. Provided the number of beneficial sites is initially large, the beneficial mutation rate will decline by a small fraction over this timescale,

$$\partial_t f_b(t + T_c) = U_b(t) - \mathcal{O}(\tilde{k}/L \cdot f_b) . \tag{317}$$

This suggests that the role of recombination can remain relatively stable over multiple fixation timescales and a large number of fixation events. Over longer timescales, as the beneficial fraction appreciably declines, the role of recombination can dramatically shift. For example, with fewer target sites, the rate of producing colocalized mutations may commensurately decline. This can lead to systematic changes in how mutations take over and minimize the impact of recombination. Provided, the number of beneficial sites is sufficiently large, $\tilde{k} \ll L \cdot f_b$. this process will occur over long timescales. As a result, we can simply extend our existing results by defining a time dependent effective mutation rate $U_b \to U_b(t)$ and substituting $U_b(t)$ into our theoretical predictions.

Diminishing returns can also arise from systematic shifts in the fitness benefits of mutations that are successful over time. For example, in any new environment, there may be a small set of large effect mutations that rapidly take over. If these beneficial sites are not replenished, the tail of the distribution

of fitness effects may decline over time. This can lead to corresponding shifts in the fitness benefits of successful mutations and the role of recombination. One simple example may arise from an initial exponential distribution of fitness benefits, where sites are replaced from the same distribution. In the quasi-asexual limit, the typical successful mutation will be drawn near the tail of the exponential $\tilde{s_b} \sim x_c$. For a set of $f_b \cdot L$ beneficial loci, this maximum sampled fitness benefit will typically be $\sim \log(f_b \cdot L)$, which for a sufficiently large number of sites will occur well above the mean. As these mutations fix, they will typically be replaced by mutations near the mean fitness benefit. This suggests that the distribution will evolve as a truncated exponential,

$$\mu(s,t) \propto \begin{cases} \frac{1}{\sigma} \cdot e^{-s/\sigma} \text{ if } s \leq c(t) \ , \\ 0 \text{ if } s > c(t) \ , \end{cases} \tag{318}$$

where the cutoff value systematically declines proportional to the substitution rate,

$$c(t) \approx \log(f_b \cdot L) - \int_0^t \theta(t')dt' \ . \tag{319}$$

Since the substitution rate scales with $1/\tilde{s_b}$, we can crudely approximate,

$$\tilde{s_b} \sim c(t) \sim \sqrt{\log(f_b \cdot L)^2 - 2t} \ . \tag{320}$$

This result allows us to leverage our single $\tilde{s_b}$ mapping to understand the role of recombination given a time-dependent distribution of fitness effects. In particular, we can substitute $\mu(s) \to \mu(s,t)$ and predict the fraction of mutations taking over on their original background.

15.2 Non-uniform distributions of beneficial loci

In this work, we have also assumed that beneficial mutations are randomly and uniformly distributed along the genome. In practice, empirical studies show that beneficial sites may be limited to a small number of 'target' or 'hotspot' genes. This concentration of beneficial sites is important because it can increase the rate at which colocalized pairs and triplets are generated. We can understand this process by considering a genome with n total genes of which a fraction $n_b \equiv f_{genes} \cdot n$ contain beneficial mutations. Each gene spans L/n base pairs and hotspot genes are spaced by a characteristic distance L/n_b. Within these hotspot genes the same total number of beneficial loci $L \cdot f_b$ are randomly and uniformly distributed. This extension leads to simple asymptotic limits.

When $f_{genes} = 1$, this model relaxes to the same uniform distribution of mutations we consider in the main text. As f_{genes} declines the probability of generating a colocalized new mutation will simply scale as,

$$\Pr[\text{new mutation in tuplet of size n}] \approx \left(\max\left(\frac{\tilde{k} \cdot \tilde{\ell}}{L f_{genes}}, \frac{\tilde{k}}{n_b} \right) \right)^{n-1} \ , \tag{321}$$

where we have assumed that the typical transfer length spans a gene $\tilde{\ell} \geq L/n$. This suggests there is a crossover $n \sim L/\tilde{\ell}$, where the probability of generating a colocalized pair will increase above the uniform limit. As f_{genes} becomes small this can lead to recombination becoming important at far smaller rates than the simple model. In extreme scenarios where the number of target genes is $\mathcal{O}(\tilde{k})$, the dynamics can become more complicated. This is because almost all mutation will arise as colocalized pairs or triplets. However, transferring these mutations across backgrounds becomes increasingly difficult since transfer events become more likely to overwrite mutations.

15.3 Pairwise epistasis

We consider a simple extension of our additive effect model that includes pairwise epistatic interactions. In particular, we assume that each mutation continues to occur at the same total rate and confer the same first order benefit s_b. We also now allow the fitness benefit to depend on the sum of pairwise fitness contributions from all other beneficial sites. As a result, this second order contribution will depend on the genotype as it evolves. For simplicity, we consider a random initialized ancestral binary genome. As a genome accrues beneficial mutations, each bit flip introduces an additional pre-determined pairwise contribution to fitness of magnitude ϵ. This allows us to write the fitness benefit of a new mutation at site i as,

$$s_i(\vec{g}) = s_b + \epsilon \cdot s_b \sum_{j \neq i} \delta_{\pm}(g_i, g_j) \ , \tag{322}$$

where $\delta_{\pm}(g_i, g_j)$ is a random and fixed 2×2 matrix of 0, $+1$ and -1 values that satisfies $\delta_{\pm}(g_j, g_i) = \delta_{\pm}(g_i, g_j)^T$. For each of the $\sim (L \times f_b)^2$ pairs, the ancestral state of $\delta_{\pm}$ is set to 0 while the sign of entries corresponding to mutated combinations is randomly sampled such that a fraction f_{pos} of are positive. This allows us to tune the average random effect of pairwise epistasis. In particular, since the number of epistatic interactions is large, we can approximate the distribution of fitness at early times as roughly constant,

$$\mu(s|\vec{g}) \approx \mu(s) \approx U_b \cdot \mathcal{N}(s_b + \epsilon s_b \cdot f_b \cdot L \cdot (2f_{pos} - 1), (2\epsilon \cdot s_b)^2 \cdot f_b \cdot L \cdot f_{pos}(1 - f_{pos})), \tag{323}$$

which simplifies in the special case $f_{pos} = 1/2$,

$$\mu(s) \approx U_b \cdot \mathcal{N}(s_b, (\epsilon \cdot s_b)^2 \cdot f_b \cdot L), \tag{324}$$

Motivated by empirical observations that epistatic interactions can be stronger when genetic loci are colocalized, we also assume that matrix values beyond $\tilde{\ell}_{epi} > |i - j|$

15.4 Uncorrelated beneficial loci

While our work assumes that the location of beneficial loci remains constant, the location of these sites may vary as the population adapts. This model is motivated by empirical evidence that particular gene knockouts can vary from beneficial to deleterious over short laboratory evolution timescales. From the perspective of a recombining population, these effects are important because they can reduce the number of background mutations can transfer to. We can begin to study these effects by considering a simple setup where a clone redraws a fraction of its beneficial loci f_{new} with every new mutation. Over moderate genetic distances, this model ensures that beneficial loci are entirely uncorrelated. We can leverage our heuristic explanations of successful mutations' evolutionary trajectories to understand how this lack of correlation diminishes the role of recombination. In the quasi-asexual regime, a successful mutation will typically transfer from the bulk to the nose before taking over by producing linked mutation. As a result, the donor and acceptor genotype will be separated by $\sim \mathcal{O}(q_a)$ mutations and the effective transfer rate will be diminished by a factor $\sim (1 - f_{new})^{q_a}$. By comparison, colocalized pairs will transfer $x_{cr}/2s_b$ times. Following each of these times the pair will be increasingly less likely to share beneficial loci with its descendants. This leads to an effective recombination rate,

$$r_{eff} \sim r \cdot \prod_{n=1}^{x_{cr}/2s_b} (1 - f_{new})^{2x_{cr}/s_b} \approx (1 - f_{new})^{\frac{x_{cr}(x_{cr}+2)}{4}} \ , \tag{325}$$

where the factor of 2 in the exponent captures the fact the both mutations in the pair must remain beneficial.

16 Numerical Methods

16.1 Forward time simulations

We validated our theoretical predictions by comparing them to forward-time, Wright-Fisher simulations. We begin with a population of $N_{t=0} \equiv N$ individuals at generation 0. Each individual is assigned a binary genome of length L. A predetermined subset of $f_b \cdot L$ loci are determined to be beneficial upon mutation from the ancestral state.

In each generation, every individual i in the population produces a Poisson number of identical offspring with mean

$$\lambda_0 = C(t) \cdot (1 + X_i - (t))[1 - \int \mu(s) ds] \,, \tag{326a}$$

and a Poisson number of mutant offspring with mean

$$\lambda_m = C(t) \cdot (1 + X_i - (t)) \int \mu(s)\, ds \,, \tag{326b}$$

where X_i is the absolute fitness of individual i, $= \sum_i X_i / \sum_i 1$ is the mean fitness of the population, and $C(t) = N/N_t$ is a normalization constant that ensures that the expected population size in each generation is equal to N. Each mutant offspring is assigned a new fitness value $X_i + s$, where s is randomly sampled from the normalized version of $\mu(s)$.

Following mutation, a Poisson number of recombination events with mean $r \cdot L \gg 1$ occur. During each recombination event, two random donor and acceptor genomes are sampled relative to their overall frequency in the population. Additionally a transfer length ℓ is randomly sampled from the distribution $\lambda(\ell)$ and a random start site i is randomly and uniformly sampled along the genome. During each transfer event, all polymorphic sites on the acceptor from locus i to $i + \ell$ are overwritten and replaced by the donor genome.

A copy of our implementation in Python is available in the associated Github repository (https://github.com/bgoodlab/TODO).

16.2 Theoretical predictions

The theory curves for each figure were generated using the following procedures.

Figure 1 The theoretical predictions for the fraction of mutations fixed on their original background in Fig. 1A,B were calculated as,

$$\text{Fraction mut. fixed on original back.} = \frac{1}{\theta_{r,>}/\theta_a - 1 \; + \theta_{r,<2}/\theta_a - 1 + \theta_{r,<3}/\theta_a - 1 + 1} \tag{327}$$

where the relative substitution rate of transferred single mutations, colocalized pairs, and colocalized triplets were calculated using Eq. (78), Eq. (95), and Eq. (114) respectively. The value of x_{ca} was calculated using a generalized version of Eq. (42),

$$1 = \frac{U_b}{s_b}\left[e^{\frac{x_{ca}s_b}{v_a}-\frac{s_b^2}{2v_a}} + \frac{\sqrt{2\pi}s_b^2}{x_{ca}\sqrt{v_a}}e^{\frac{x_{ca}^2}{2v_a}}\Phi\left(\frac{s_b - x_{ca}}{\sqrt{v_a}}\right)\right] . \tag{328}$$

Numerical solutions to Eq. (328) were obtained using the `fsolve` function from the SciPy library (58), with the measured value of v_a obtained from our forward-time simulations. Solutions for $x_{cr2<}$ and $x_{cr3<}$ were also obtained using Eq. (328) with the substitutions $U_b \to r\cdot\tilde{\ell}/2$ and $s_b \to 2s_b$ for colocalized pairs and $U_b \to r\cdot\tilde{\ell}/2$ and $s_b \to 3s_b$. The theoretical predictions for the rate of adaptation were obtained by numerically solving,

$$1 \approx \frac{2NU_b x_{ca}}{\sqrt{2\pi v}}\cdot e^{-\frac{(x_{ca}-s_b)^2}{2v}}\left[1 + \frac{r\cdot\tilde{\ell}\cdot\tilde{k}}{U_b} + \frac{\tilde{\ell}\cdot\tilde{k}}{L}\cdot\frac{x_{cr2}}{x_{ca}}e^{\frac{(x_{ca}-s_b)^2}{2v}-\frac{(x_{cr2}-s_b)^2}{2v}} + \left(\frac{\tilde{\ell}\cdot\tilde{k}}{L}\right)^2\cdot\frac{x_{cr3}}{x_{ca}}e^{\frac{(x_{ca}-s_b)^2}{2v}-\frac{(x_{cr3}-s_b)^2}{2v}}\right] . \tag{329}$$

Figure 3 The theoretical prediction for the asexual distance between consecutively fixed mutations in Fig. 3B is simply a truncated exponential distribution over the interval $(0, L)$ with mean $L/\tilde{k}$,

$$\Pr_{\text{asexual}}(\delta\ell_2) \equiv \frac{\frac{\tilde{k}}{L}\cdot e^{-\tilde{k}/L\cdot\delta\ell_2}}{1-e^{-\tilde{k}}} . \tag{330}$$

For the simulated parameters, contributions to the sexual distribution will predominantly arise from mutations fixing asexually and colocalized pairs. The list of fixed polymorphisms will thus be a sequence of colocalized pairs separated by a distance $\delta\ell_2 < \tilde{\ell}$ and random integers sampled from Eq. (330). It follows that the theoretical prediction for the sexual distribution is an exponential distribution weighted by relative fixation probability,

$$\Pr_{\text{sexual}}(\delta\ell_2) \equiv \frac{\Pr_{\text{asexual}}(\delta\ell_2)\cdot\left[\zeta\cdot\frac{x_{cr}^2(\delta\ell_2)}{x_c s_b}\cdot e^{\frac{(x_c-s_b)^2}{2v}-\frac{(x_{cr}(\delta\ell_2)-s_b)^2}{2v}}\cdot\delta(\delta\ell_2<\tilde{\ell}) + (1-\zeta)\cdot 1\cdot\delta(\delta\ell_2\geq\tilde{\ell})\right]}{\int_0^L \delta\ell_2\,\Pr_{\text{asexual}}(\delta\ell_2)\cdot\left[\zeta\cdot\frac{x_{cr}^2(\delta\ell_2)}{x_c s_b}\cdot e^{\frac{(x_c-s_b)^2}{2v}-\frac{(x_{cr}(\delta\ell_2)-s_b)^2}{2v}}\cdot\delta(\delta\ell_2<\tilde{\ell}) + (1-\zeta)\cdot 1\cdot\delta(\delta\ell_2\geq\tilde{\ell})\right]} , \tag{331}$$

where $\delta(x>y)$ is equal to 1 if $x>y$ and 0 otherwise and ζ accounts for the probability that the sampled distance involved both mutations in the colocalized pair,

$$\zeta = \frac{1-\frac{\theta_a}{\theta_{r2<}}}{2-\frac{\theta_a}{\theta_{r2<}}} . \tag{332}$$

The theoretical prediction for the cumulative density function of distances between segregating mutations $\delta\ell$ in Fig. 3C is simply obtained by integrating Eq. (330),

$$\Pr(\delta\ell < \delta\ell')\int_0^{\delta\ell'}\Pr_{\text{asexual}}(\delta\ell)\,d\delta\ell. \tag{333}$$

The theoretical prediction for the fixation probability in figFig. 3D is,

$$p_{fix}(\delta\ell_2) = \begin{cases} \frac{x_{cr}^2(\delta\ell_2)}{x_c s_b} \cdot e^{\frac{(x_c - s_b)^2}{2v_a} - \frac{(x_{cr}(\delta\ell_2) - s_b)^2}{2v_a}} \text{ if } \delta\ell_2 < \tilde{\ell} \ , \\ 1 \text{ else} \ . \end{cases} \tag{334}$$

The theoretical predictions from the phase diagram were obtained from Eq. (104), Eq. (106), Eq. (125), and Eq. (127).