Linkage disequilibrium between rare mutations

Benjamin H. Good[1]

[1] *Department of Applied Physics, Stanford University, Stanford, CA 94305*

The statistical associations between mutations, collectively known as linkage disequilibrium (LD), encode important information about the evolutionary forces acting within a population. Yet in contrast to single-site analogues like the site frequency spectrum, our theoretical understanding of linkage disequilibrium remains limited. In particular, little is currently known about how mutations with different ages and fitness costs contribute to expected patterns of LD, even in simple settings where recombination and genetic drift are the major evolutionary forces. Here, I introduce a forward-time framework for predicting linkage disequilibrium between pairs of neutral and deleterious mutations as a function of their present-day frequencies. I show that the dynamics of linkage disequilibrium become much simpler in the limit that mutations are rare, where they admit a simple heuristic picture based on the trajectories of the underlying lineages. I use this approach to derive analytical expressions for a family of frequency-weighted LD statistics as a function of the recombination rate, the frequency scale, and the additive and epistatic fitness costs of the mutations. I find that the frequency scale can have a dramatic impact on the shapes of the resulting LD curves, reflecting the broad range of time scales over which these correlations arise. I also show that the differences between neutral and deleterious LD are not purely driven by differences in their mutation frequencies, and can instead display qualitative features that are reminiscent of epistasis. I conclude by discussing the implications of these results for recent LD measurements in bacteria. This forward-time approach may provide a useful framework for predicting linkage disequilibrium across a range of evolutionary scenarios.

The statistical associations between mutations, collectively known as linkage disequilibrium (LD), play a central role in modern evolutionary genetics (Slatkin, 2008). Correlations between mutations enable genome-wide association studies and related methods for mapping genetic basis of diseases and other complex traits (Visscher et al., 2017). The co-occurrence of mutations in specific DNA molecules is also important for evolutionary dynamics, since these realized combinations provide the raw material on which natural selection and other evolutionary forces can act. As a result, contemporary patterns of linkage disequilibrium encode crucial information about the historical processes of recombination (McVean et al., 2004; Rosen et al., 2015), natural selection (Garud et al., 2015; Sabeti et al., 2002), and demography (Harris and Nielsen, 2013; Li and Durbin, 2011; Ragsdale and Gravel, 2019) that operate within a population. Yet while extensive theory has been developed for predicting marginal distributions of mutations (Coop and Ralph, 2012; Cvijović et al., 2018; Kamm et al., 2017; Kimura, 1964; Neher and Hallatschek, 2013; Polanski and Kimmel, 2003; Sawyer and Hartl, 1992), higher order correlations like linkage disequilibrium remain poorly understood in comparison.

Many studies of linkage disequilibrium have focused on pairwise correlations between mutations at different sites along a genome. These correlations are often summarized by the correlation coefficient,

$$r^2 = \frac{(f_{AB} - f_A f_B)^2}{f_A(1-f_A) f_B(1-f_B)} \,, \tag{1}$$

where f_{AB} is the fraction of individuals in the population with mutations at both sites, and f_A and f_B are the marginal frequencies of the two mutations (Hill and Robertson, 1968). The r^2 statistic and related measures like D' (Lewontin, 1964) quantify how the joint distribution of the two mutations differs from a null model in which the alleles are independently distributed across individuals. A celebrated theoretical result by Ohta and Kimura (1971) shows that, for a neutrally evolving population of constant size N, the frequency-weighted expectation of r^2 is given by

$$\mathbb{E}[r^2 \cdot W(f_A, f_B)] = \frac{5 + NR}{11 + 13NR + 2(NR)^2} \,, \tag{2}$$

where R is the recombination rate between the two sites, and $W(f_A, f_B) \propto f_A(1-f_A) f_B(1-f_B)$ is a weighting function that is normalized so that $\mathbb{E}[W(f_A, f_B)] = 1$. The expression in Eq. (2) approaches an $\mathcal{O}(1)$ constant in the limit of low recombination rates ($NR \ll 1$) and decays as $\sim 1/NR$ when $NR \gg 1$. Similar behavior also occurs for the unweighted expectation $\mathbb{E}[r^2]$ (McVean, 2002), which shares the same $\sim 1/NR$ scaling when $NR \to \infty$ (Song and Song, 2007). As a result, the shape of this LD curve is frequently used to estimate rates of recombination, e.g. by examining how genome-wide averages of linkage disequilibrium decay as a function of the coordinate distance between sites (Ansari and Didelot, 2014; Chakravarti et al., 1984; Garud et al., 2019; Lynch et al., 2014; Rosen et al., 2015).

While these classical results have been enormously influential for building intuition about linkage disequilibrium, they suffer from several limitations that are increasingly important in modern genomic datasets. Chief among these is the absence of natural selection. While there has been some progress in predicting patterns of linkage disequilibrium under particular selection scenarios [e.g., hitchhiking near a recent selective sweep (Kim and Nielsen, 2004; McVean, 2007; Pfaffelhuber et al., 2008; Pokalyuk, 2012; Stephan et al., 2006)] we currently lack analogous theoretical predictions for the empirically relevant case where a subset of the observed mutations are deleterious. This is a crucial limitation, since numerous studies have documented differences in the genome-wide patterns of LD between synonymous and nonsynonymous mutations (Arnold et al., 2020; Garcia and Lohmueller, 2020; Rosen et al., 2018; Sohail et al., 2017) or for genic vs intergenic regions of the genome (Eberle et al., 2006), where purifying selection is thought to play an important role. Several recent studies have begun to explore these effects in computer simulations (Arnold et al., 2020; Garcia and Lohmueller, 2020; Ragsdale and Gravel, 2019). Yet without a corresponding analytical theory, it can be difficult to understand how these patterns depend on the underlying parameters of the model, or to determine when more exotic forces like positive selection, epistasis, or ecological structure are necessary to fully explain the observed data.

A second and related limitation arises from the averaging scheme in Eq. (2), which weights each pair of mutations by their joint heterozygosity, $W(f_A, f_B) \propto f_A(1-f_A)f_B(1-f_B)$. This weighting tends to favor mutations with intermediate frequencies (e.g., $10\% \leq f \leq 90\%$); these mutations are represent older genetic variants that have been segregating for times comparable to the most recent common ancestor of the population (McVean, 2002; Rogers, 2014). Yet in practice, even a single population will typically harbor mutations across an enormous range of frequency scales, reflecting the broad range of timescales at which these mutations occurred (Cvijović et al., 2018). This broad range of frequencies is increasingly accessible in modern genomic datasets, where sample sizes can range from several hundred to several hundred thousand individuals (Allix-Béguec et al., 2018; Karczewski et al., 2020; Pasolli et al., 2019; Petit III and Read, 2018; Shu and McCauley, 2017). Understanding how LD varies across these different frequency scales could therefore provide new information about the evolutionary processes that operate on different ancestral time scales, similar to existing approaches based on the single-site frequency spectrum (Lawrie and Petrov, 2014; Ragsdale et al., 2018). Such an approach could be particularly useful for probing aspects of the recombination process, which are difficult to observe from single-site statistics alone.

Yet at present, little is known about how different frequency and time scales contribute to the expected patterns of linkage disequilibrium within a population. Previous theoretical work has explored how mutation frequencies constrain the possible values of statistics like r^2, independent of the underlying evolutionary dynamics (Hedrick, 1987; Kang and Rosenberg, 2019; Lewontin, 1988; VanLiere and Rosenberg, 2008). However, few methods currently exist for predicting the quantitative values that are expected to emerge under a given evolutionary scenario. This limited frequency resolution is particularly problematic in the presence of natural selection, which is known to strongly influence the distribution of mutation frequencies. This makes it difficult to interpret the varying LD patterns that have been observed across different classes of selected sites in a variety of natural populations. Do the differences between synonymous and nonsynonymous LD arise purely due to differences in their mutation frequencies? Or are there residual signatures of selection that remain even after controlling for marginal mutation frequencies? What conclusions can we draw about the underlying selection and recombination processes when differences are observed for some frequency ranges but not others? The goal of this work is to develop the theoretical tools necessary to address these questions.

In the following sections, I present an analytical framework for predicting linkage disequilibrium between pairs of neutral and deleterious mutations as a function of their present-day frequencies. I do so by generalizing the traditional weighted average in Eq. (2), defining a family of weights $W(f_A, f_B|f_0)$ that preferentially exclude mutations with frequencies $\gtrsim f_0$. I show that the dynamics of linkage disequilibrium become much simpler in the limit that mutations are rare ($f_0 \ll 1$), where they can be analyzed using a forward-time branching process framework. I use this approach to derive analytical expressions for statistics like r^2 as a function of the recombination rate, the frequency scale f_0, and the additive and epistatic fitness costs of the two mutations. I find that the frequency scale f_0 can have a dramatic impact the shape of the weighted LD curve, reflecting the varying timescales over which these mutations persist within the population. I show how this scaling behavior can serve as a probe for estimating recombination rates and distributions of fitness effects in large cohorts, and I discuss the implications of these results for recent LD measurements in bacteria. This forward-time approach may provide a useful framework for predicting linkage disequilibrium across a range of other evolutionary scenarios.

Data Availability

Source code for forward-time simulations, data analysis, and figure generation are available at Github (`https://github.com/bgoodlab/rare_ld`). Polymorphism data from *E. rectale* were obtained from a previous study (Garud et al. (2019)) and can be accessed using the accessions listed in that work. Postprocessed SNV data have been deposited in the GSA Figshare portal.

MODEL AND RESULTS

Here we investigate the dynamics of linkage disequilibrium between a pair of genomic loci in a population of constant size N. We assume that each locus accumulates mutations at rate μ per individual per generation, and we focus on the infinite sites limit where $N\mu \ll 1$. The mutant and wildtype alleles at each locus are denoted by A/a and B/b, respectively. We assume that mutations at these loci reduce the fitness of wildtype individuals by an amount s_A and s_B, respectively, while the fitness of the double mutant is reduced by $s_{AB} = s_A + s_B + \epsilon$. The parameter ϵ allows us to account for epistatic interactions between the two mutations, while the additive limit is recovered when $\epsilon = 0$. Since we envision eventual applications to bacteria, we will primarily focus on haploid genomes where we can neglect the further complications of dominance and ploidy. However, our main results will also apply to diploid organisms when mutations are semidominant ($h = 1/2$).

We assume that the two loci undergo recombination at rate R per individual per generation, producing double mutant combinations from single mutants, and vice versa. These recombination events could be implemented through a variety of mechanisms, including crossover recombination, gene conversion, and homologous recombination of horizontally transferred DNA. In the context of our two-locus model, the differences between these mechanisms can be entirely absorbed in the definition of R, and will primarily influence how R scales as a function of the coordinate distance ℓ between the two loci. In simple cases, the scaling at short distances can often be captured by the linear relationship, $R(\ell) = r\ell$, where r is a measure of the recombination rate per base pair. However, all of our results will be independent of the functional form of $R(\ell)$, provided that all distances are expressed in units of map length (R). We will revisit this point in more detail when we discuss applications to genomic data below.

These assumptions yield a standard two-locus model for the population frequencies of the four possible combinations (or *haplotypes*), f_{ab}, f_{Ab}, f_{aB}, and f_{AB}, as well as the corresponding mutation (or *allele*) frequencies, $f_A \equiv f_{Ab} + f_{AB}$ and $f_B \equiv f_{aB} + f_{AB}$. In the diffusion limit (Ewens, 2004), this model can be expressed as a coupled system of nonlinear stochastic differential equations,

$$\frac{\partial f_{ab}}{\partial t} = -[0 - \overline{S}(t)]f_{ab} - RD(t) + \frac{\xi_{ab}(t)}{\sqrt{N}} + \mu(f_{Ab} + f_{aB}) - 2\mu f_{ab} \, , \tag{3a}$$

$$\frac{\partial f_{Ab}}{\partial t} = -[s_A - \overline{S}(t)]f_{Ab} + RD(t) + \frac{\xi_{Ab}(t)}{\sqrt{N}} + \mu(f_{ab} + f_{AB}) - 2\mu f_{Ab} \, , \tag{3b}$$

$$\frac{\partial f_{aB}}{\partial t} = -[s_B - \overline{S}(t)]f_{aB} + RD(t) + \frac{\xi_{aB}(t)}{\sqrt{N}} + \mu(f_{ab} + f_{AB}) - 2\mu f_{aB} \, , \tag{3c}$$

$$\frac{\partial f_{AB}}{\partial t} = -[s_{AB} - \overline{S}(t)]f_{AB} - RD(t) + \frac{\xi_{AB}(t)}{\sqrt{N}} + \mu(f_{Ab} + f_{aB}) - 2\mu f_{AB} \, , \tag{3d}$$

where $\overline{S}(t) \equiv s_A f_{Ab} + s_B f_{aB} + s_{AB} f_{AB}$ is the mean fitness reduction within the population,

$$D(t) \equiv f_{AB} f_{ab} - f_{Ab} f_{aB} = f_{AB} - f_A f_B \tag{3e}$$

is the standard coefficient of linkage disequilibrium, and $\{\xi_i(t)\}$ are a collection of Brownian noise terms with a covariance structure that depends on $\{f_i(t)\}$ (Good and Desai, 2013). This Langevin formulation is formally equivalent to the diffusion limit of population genetics, which is more commonly expressed using the Fokker-Planck equation for the probability density, $p(f_{AB}, f_{Ab}, f_{aB}, f_{ab})$ (Ewens, 2004). In this case, we will see that the Langevin notation in Eq. (3) will be slightly more convenient for our analytical calculations below. To streamline notation, we have also assumed that forward and reverse mutations occur at the same rate μ for both sites — this will not turn out to be a crucial distinction, since the mutation rates will largely cancel out when we focus on the infinite sites limit ($N\mu \to 0$).

Following previous work, our analysis will focus on measures of linkage disequilibrium that are derived from different moments of D, f_A, and f_B. For example, the weighted average in Eq. (2) is equivalent to the ratio of expectations,

$$\mathbb{E}[r^2 \cdot W(f_A, f_B)] = \frac{\langle D^2 \rangle}{\langle f_A(1-f_A) f_B(1-f_B) \rangle} \, , \tag{4}$$

where the angle brackets $\langle \cdot \rangle$ denote the expectation $\mathbb{E}[\cdot]$. This ratio of expectations is traditionally denoted by the symbol σ_d^2 (Ohta and Kimura, 1971). This quantity has the convenient property that it is independent of the mutation rate μ in the limit that $N\mu \ll 1$, which eliminates the dependence on one of the parameters. In other words, σ_d^2 is primarily measuring properties of the segregating mutations, rather than the target size for those mutations to occur. Here, we generalize this definition of σ_d^2 to consider a family of weighted averages of the form

$$\sigma_d^2(f_0) \equiv \frac{\left\langle D^2 \cdot e^{-f_A/f_0 - f_B/f_0} \right\rangle}{\left\langle f_A(1-f_A) f_B(1-f_B) \cdot e^{-f_A/f_0 - f_B/f_0} \right\rangle} \, , \tag{5}$$

where f_0 is a characteristic frequency scale. This is equivalent to choosing a weighting function,

$$W(f_A, f_B | f_0) \propto f_A(1-f_A) e^{-f_A/f_0} \times f_B(1-f_B) e^{-f_B/f_0} \, , \tag{6}$$

in the weighted expectation $\mathbb{E}[r^2 \cdot W(f_A, f_B|f_0)]$, where the constant of proportionality is again chosen such that $\mathbb{E}[W(f_A, f_B|f_0)] = 1$. The additional exponential factors in this weighting term act like a smeared out version of a step function, preferentially excluding contributions from mutations with frequencies much larger than f_0. By scanning over different values of f_0, this weighted version of $\sigma_d(f_0)$ allows us to quantify how linkage disequilibrium varies over different frequency scales. This weighting scheme is reminiscent of the frequency thresholds that have previously been employed in empirical and computational studies of LD. From a qualitative perspective, the precise shape of the weighting function will turn out to be relatively unimportant — we argue below that any sufficiently sharp transition will produce qualitatively similar behavior for $\sigma_d^2(f_0)$. However, the exponential function will turn out to have some particularly convenient properties that will be useful for our analytical calculations below.

In addition to $\sigma_d^2(f_0)$, it will also be useful to consider more general weighted moments of D defined by

$$\sigma_d^k(f_0) \equiv \frac{\left\langle D^k \cdot e^{-\frac{f_A+f_B}{f_0}} \right\rangle}{\left\langle f_A(1-f_A)f_B(1-f_B) \cdot e^{-\frac{f_A+f_B}{f_0}} \right\rangle} \,. \tag{7}$$

These moments will also be independent of μ in the infinite sites limit ($N\mu \ll 1$), so that they also capture properties of the segregating mutations. We will be particularly interested in the $k = 1$ moment, which measures the degree to which mutations are in *coupling linkage* (AB/ab haplotypes) vs *repulsion linkage* (Ab/aB haplotypes), as well as the $k = 4$ moment, which can be combined with σ_d^2 to quantify fluctuations in D^2.

Despite the simplicity of the two locus model in Eq. (3), the resulting patterns of linkage disequilibrium have been difficult to characterize theoretically. As in many population genetic problems, the major hurdles arise from nonlinearities in the $\overline{S}$, D, and $\{\xi_i\}$ terms in the stochastic differential equations, which typically require numerical approaches to make further progress. However, dramatic simplifications arise if we restrict our attention to frequency scales $f_0 \ll 1$. In this case, the exponential weights in Eq. (7) will be vanishingly small except in cases where $f_{AB},\ f_{Ab}, f_{aB} \lesssim f_0 \ll 1$, such that the vast majority of the population is comprised of wildtype individuals. Applying this approximation to the two-locus model in Eq. (3), we obtain a *nearly* linear system of stochastic differential equations for the three mutant haplotypes,

$$\frac{\partial f_{Ab}}{\partial t} = -s_A f_{Ab} + \mu + RD(t) + \sqrt{\frac{f_{Ab}}{N}}\eta_{Ab}(t) \,, \tag{8a}$$

$$\frac{\partial f_{aB}}{\partial t} = -s_B f_{aB} + \mu + RD(t) + \sqrt{\frac{f_{aB}}{N}}\eta_{aB}(t) \,, \tag{8b}$$

$$\begin{aligned}\frac{\partial f_{AB}}{\partial t} = -s_{AB} f_{AB} - RD(t) + \sqrt{\frac{f_{AB}}{N}}\eta_{AB}(t) \\ + \mu(f_{Ab} + f_{aB}) \,,\end{aligned} \tag{8c}$$

where the $\{\eta_i(t)\}$ are independent Brownian noise terms with mean zero and variance one (Gardiner, 1985). Note that there is some subtlety involved in this approximation, since our weighting scheme only places limits on the haplotype frequencies at the time of observation. Equation (8) requires the stronger assumption that $f_{AB}, f_{Ab}, f_{aB} \ll 1$ for all previous times as well. This distinction will become important in our analysis below. When the approximations in Eq. (8) hold, the only remaining nonlinearity enters through the $f_{Ab}f_{aB}$ term in $D(t)$. This term is crucial for allowing double mutants to be produced from single mutants via recombination. However, in many cases of interest, the contribution from this term will turn out to be small, either because R itself is sufficiently small, or because $D(t)$ is small (e.g. for sufficiently large R). This suggests that in many cases, we may be able to treat the nonlinearity in $D(t)$ as a perturbative correction to the otherwise linear dynamics in Eq. (8). Extensive theory has been developed for analyzing linear stochastic differential equations of this form, ranging from powerful heuristic approaches to exact analytical results (Cvijović et al., 2018; Desai and Fisher, 2007; Fisher, 2007; Good, 2016; Weissman et al., 2010, 2009). The goal of the following sections is to flesh out this basic intuition and show how it can be used to obtain predictions for linkage disequilibrium statistics like $\sigma_d^k(f_0)$. We will begin by presenting a heuristic analysis of the problem, which will allow us to identify the key timescales and dynamical processes involved, and will be useful for building intuition for the formal analysis that follows. We will conclude by discussing potential applications of these results in the context of genomic data.

HEURISTIC ANALYSIS

We begin by presenting a heuristic analysis of the two-locus model, which focuses on the underlying lineage dynamics that contribute to LD statistics like $\sigma_d(f_0)$. Roughly speaking, this heuristic analysis will be accurate to leading order in the logarithms of various quantities (e.g. mutation frequencies, recombination rates), while providing a more mechanistic picture of the underlying lineage dynamics that are involved. Readers who prefer a more formal treatment may find it useful to start with the ***Formal Analysis*** section below. Our heuristic approach is similar to the one developed by Weissman et al. (2010) to study the process of fitness valley crossing in sexual populations. A key difference in this work is that we are now more interested in understanding the steady-state frequency distributions of new mutations, rather than the transient process of valley crossing.

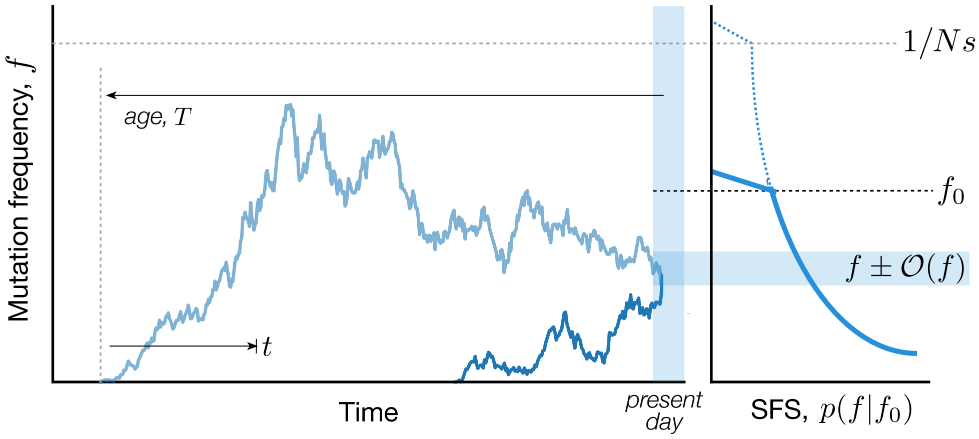

FIG. 1 Schematic illustration of mutation trajectories that contribute to the site frequency spectrum. Left: Mutations arise at different times and drift to their present-day frequencies (shaded region). Dark and light blue lines show examples of present-day mutations with *upward* and *downward* trajectories, respectively. In both cases, deleterious mutations are prevented from growing much larger than the drift barrier, $f_{\rm sel}{\sim}1/Ns$ (grey dashed line). Right: The site frequency spectrum (SFS) is the sum of the probabilities of all mutation trajectories with a present-day frequency f. Each mutation can be characterized by its age T and historical trajectory $f(t)$, with $f(T) \sim f$. When the frequency spectrum is dominated by upward trajectories, the effects of negative selection are similar to imposing a present-day frequency threshold f_0 (black dashed line).

Single-locus dynamics

Before considering the full two-locus problem, it will be useful to briefly review the evolutionary dynamics that arise at a single genetic locus. We will focus on the *infinite sites limit* ($N\mu \ll 1$) where the population is almost always composed of wildtype individuals, and new mutations are introduced into the population at a total rate $N\mu \ll 1$ per generation. The vast majority of these lineages will drift to extinction without ever leaving more than a few descendants in the population. However, a lucky minority will survive for sufficiently long times that they can grow to reach observable frequencies. The dynamics of this process will strongly depend on the fitness cost s of the mutation. We consider both neutral ($s = 0$) and deleterious ($s > 0$) mutations below.

Neutral mutations. When $s = 0$, the frequency of a mutant lineage is only influenced by genetic drift. These dynamics take a particularly simple form when the mutation is sufficiently rare [$f(t) \ll 1$], since the sub-lineages founded by different mutant individuals will evolve independently of each other. With probability $\sim 1/t$, these individual mutant lineages will survive for at least $\sim t$ generations and reach a characteristic size $f(t){\sim}t/N$ (Fisher, 2007). More detailed calculations (**?**) show that the frequency of the surviving lineage is exponentially distributed around this characteristic size, so that

$$p(f|t, f{>}0)\, df \sim \frac{N}{t} e^{-Nf/t}\, df\,. \tag{9}$$

Most of this distribution is concentrated within an order of magnitude of the mean ($\sim t/N$), consistent with our notion of a "typical" value. However, there is also a small probability ($\sim Nf/t$) for the mutation to be observed at much lower frequencies ($0{<}f{\ll}t/N$). This will become important in our discussion below.

Essentially all of the results in this work will follow from repeated applications of these basic temporal dynamics. For example, the total probability of observing a mutation with a present-day frequency f can be calculated by summing Eq. (9) over the possible times that the mutation could have originated in the past (Fig. 1). This allows us to recover the familiar neutral site frequency spectrum,

$$p(f)df \sim \int \underbrace{N\mu\, dT}_{\text{occurs}} \cdot \underbrace{\left(\frac{1}{T}\right)}_{\text{survives}} \cdot \underbrace{\frac{N}{T} e^{-\frac{Nf}{T}}\, df}_{p(f|T,f>0)df} \sim \frac{N\mu}{f}\, df\,. \tag{10}$$

We can impose a maximum frequency threshold $\sim f_0$ by inserting a unit step function, $u(f - f_0)$, so that

$$p(f|f_0) \sim \frac{N\mu \cdot u(f_0 - f)}{f} \equiv \begin{cases} \frac{N\mu}{f} & \text{if } f < f_0, \\ 0 & \text{else.} \end{cases} \tag{11}$$

In contrast to the transient distribution in Eq. (9) — which was concentrated around a typical frequency $\sim t/N$ — the equilibrium site frequency spectrum in Eq. (11) is evenly distributed on a *log scale*, with equal contributions from all orders of magnitude between 0 and f_0. Since there are many more orders of magnitude near 0 than near f_0, this implies that most of the probability will be concentrated at extremely low frequencies — so much so that the distribution in Eq. (11) is not even normalizable. However, these extremely rare mutations will only have a negligible influence on general moments of the form

$$\langle f^p | f_0 \rangle \sim \int_0^{f_0} f^p \cdot \frac{N\mu}{f}\, df \sim \frac{N\mu f_0^p}{p}\,, \tag{12}$$

which are dominated by mutations with frequencies of order $\sim f_0$ when $p \gtrsim 1$. Similarly, the total probability of observing a mutation with a frequency of order $\sim f_0$ has a constant value,

$$\Pr[f \sim f_0] \equiv \int_{f\sim f_0} \frac{N\mu}{f}\, df \sim N\mu\,, \tag{13}$$

where the integral denotes a sum over frequencies in the range $\log f = \log f_0 \pm \mathcal{O}(1)$. These integrated probabilities are insensitive to the divergence at low frequencies, and will play a crucial role in our analysis below.

Historical trajectories of mutations. Our linkage disequilibrium analysis will also require information about the historical trajectories of mutations that are sampled at a given frequency in the present. We can analyze these trajectories using the same set of tools that we used to derive the site frequency spectrum above. For example, the sum in Eq. (10) suggests that the age (T) of a mutation with present-day frequency f (Fig. 1) will follow an inverse exponential distribution,

$$p(T|f)dT \sim \frac{Nf}{T^2} e^{-Nf/T}\, dT\,, \tag{14}$$

which has a peak at $T{\sim}Nf$ and a broad power law tail for $T{\gg}Nf$. These different scalings suggest that it will be useful to consider the trajectories of mutations with ages $T{\sim}Nf$ and $T{\gg}Nf$, respectively.

Mutations in the first category ($T{\sim}Nf$) will tend to have "upward trajectories" (Fig. 1) similar to the unconstrained dynamics in Eq. (9). We can see this by considering the frequency of the mutation at some intermediate timepoint t (Fig. 1) and expressing the present-day frequency as a sum over the $Nf(t)$ individual lineages that were founded at time t. The conditional distribution of $f(t)$ then follows from Bayes rule,

$$p(f(t)|f,T) \propto p\left(\tfrac{1}{N} {\to} f(t)|t\right) \cdot \Pr\left[\textstyle\sum_{i=1}^{Nf(t)} f_i(T){\sim} f\right]\,, \tag{15}$$

where $p(1/N{\to}f(t)|t)$ is the prior probability of transitioning from $1/N$ to $f(t)$, and $f_i(T)$ is the present-day frequency of the ith intermediate lineage. From our discussion above, each of these intermediate lineages has a probability $\sim 1/(T{-}t)$ of surviving until time T and growing to size $f_i(T){\sim}(T{-}t)/N$. When $T{-}t \ll T$, many independent lineages will survive to contribute to the present-day frequency $f(T){\sim}f$, and $f(t)$ will remain close to f by the central limit theorem. On the other hand, when $t \ll T$, the typical size of a single surviving lineage ($f_i(T){\sim}T/N$) will already be comparable to the total present-day frequency ($f(T){\sim}f$). This implies that the most likely way for the entire mutation to transition from $f(t)$ to $\mathcal{O}(f)$ is for all but one of the intermediate lineages to go extinct. The total probability of this coarse-grained trajectory can therefore be approximated by

$$p(f(t)|f,T) \propto \underbrace{\frac{1}{t} \times \frac{N}{t} e^{-\frac{Nf(t)}{t}}}_{p(1/N \to f(t)|t)} \times \underbrace{\frac{Nf(t)}{T} e^{-\frac{Nf(t)}{T}}}_{\text{all but one extinct}}\,, \tag{16}$$

which reduces to

$$p(f(t)|f,T) \propto f(t) e^{-Nf(t)/t} \tag{17}$$

in the limit that $t \ll T \sim Nf$. This conditional distribution has a typical frequency of order $f(t){\sim}t/N$, similar to the unconditioned distribution in Eq. (9). However, the historical trajectories have a much smaller probability of drifting to anomalously low frequencies at the intermediate timepoints ($f(t){\ll}t/N{\ll}f$), since they have been conditioned on reaching much higher frequencies in the present ($f(T){\sim}f$).

In contrast to these upward trajectories, older mutations ($T{\gg}Nf$) will tend to have "downward trajectories" similar to the one depicted in Fig. 1 (light blue line). These downward trajectories can be quantified using a similar lineage-based picture as above. The major difference is that there is now a broad range of intermediate timepoints ($T{-}t{\gg}Nf$) where the size of a single surviving lineage $[f_i(T){\sim}(T-t)/N]$ is much larger than the final frequency $f(T){\sim}f$. In order to transition from $f(t)$ to f, the sole surviving lineage must now also drift to an anomalously low frequency by the time of observation, while simultaneously avoiding extinction. Equation (9) shows that the probability of such an excursion is of order $\sim Nf/(T-t)$, so that the total probability of transitioning from $f(t)$ to f is given by

$$\Pr[\textstyle\sum_i f_i(T){\sim}f] \sim \underbrace{\frac{Nf(t)}{T-t} e^{-\frac{Nf(t)}{(T-t)}}}_{\text{all but one extinct}} \times \underbrace{\frac{Nf}{T-t}}_{\Pr[f_i(T)\sim f]}\,. \tag{18}$$

The conditional distribution of $f(t)$ then reduces to

$$p(f(t)|f,T) \propto f(t) e^{-NTf(t)/t(T-t)}\,, \tag{19}$$

which is valid in the limit that T and $T{-}t$ are both much larger than Nf.

In contrast to the upward trajectories in Eq. (17), the typical frequencies in Eq. (19) now have a non-monotonic dependence on the intermediate timepoint t,

$$f(t) \sim \frac{t(T-t)}{NT}\,. \tag{20}$$

These historical frequencies attain their maximum value near the middle of the trajectory ($t{\approx}T/2$) and decrease toward the present day. This motivates our "downward trajectory" naming scheme, since these older mutations will typically be decreasing in frequency at the time that they are sampled (Fig. 1). When $T{\gg}Nf$, these historical frequencies can be much larger than any present-day frequency threshold f_0, and can even reach a point where the rare mutation assumption starts to break down [$f(t){\sim}1$]. Fortunately, Eq. (10) shows that these ancient mutations will have a negligible influence on the site frequency spectrum, which is dominated by contributions from upward trajectories with $T{\sim}Nf$.

However, downward trajectories can still play an important role in other evolutionary quantities, even when they provide a negligible contribution to the site frequency spectrum itself. One of the simplest examples is the average age of a mutation with a present-day frequency f. Although the *median age* in Eq. (14) occurs for $T{\sim}Nf$ — which provides the dominant contribution to the site frequency spectrum — the $\sim 1/T^2$ tail in Eq. (14) causes the *average age* to be dominated by increasingly older mutations with $T{\gg}Nf$. This cannot continue indefinitely, however, since the historical sizes of these mutations will eventually become large enough that our rare mutation assumption will break down. For example, ancient mutations that drift to $\mathcal{O}(1)$ frequencies will eventually be more likely to fix than to drift back down to $f(T){\sim}f$ by the time of observation. We can account for this behavior in a crude way by truncating the integral in Eq. (10) at a maximum age $T_{\max}{\sim}N$, which corresponds to a maximum historical frequency $f_{\max}{\sim}1$. This yields a finite value for the average age of a mutation as a function of its present-day frequency,

$$\langle T|f\rangle \sim \int_0^{T_{\max}} T\cdot\frac{Nf}{T^2}e^{-\frac{Nf}{T}}\,dT \sim Nf\log\left(1/f\right)\,, \quad (21)$$

which matches the well-known result from Kimura and Ohta (1973) in the limit that $f \ll 1$. This same cutoff approximation will play a crucial role in analyzing the dynamics of linkage disequilibrium below.

Deleterious mutations. Similar considerations apply for deleterious mutations ($s > 0$), except that natural selection will prevent them from growing much larger than a critical frequency, $f_{\rm sel}{\sim}1/Ns$, above which natural selection starts to dominate over genetic drift (Fisher, 2007; Good, 2016). Conversely, genetic drift will continue to dominate over natural selection for frequencies $f(t) \ll f_{\rm sel}$, and the dynamics will resemble those derived for neutral mutations above. This suggests that we can approximate the deleterious site frequency spectrum simply by adding an additional step function to the neutral result,

$$p(f|f_0) \sim \frac{N\mu\cdot u(f_0-f)\cdot u\left(\frac{1}{Ns}-f\right)}{f}\,, \quad (22)$$

which enforces the condition that mutations will rarely be found above the"drift barrier," $f_{\rm sel}{\sim}1/Ns$.

The net effect of this new threshold will depend on the compound parameter Nsf_0, which captures the relative strength of selection on timescales of order $\sim Nf_0$. When $Nsf_0 \ll 1$, deleterious mutations are always sampled in their effectively neutral phase, and the site frequency spectrum reduces to the neutral version in Eq. (11). On the other hand, when $Nsf_0 \gg 1$, the frequency spectrum maintains a similar shape, but with an effective frequency threshold now set by the drift barrier $f_{\rm sel}{\sim}1/Ns \ll f_0$. This implies that averages like $\langle f^p\rangle$ will be dominated by frequencies of order $\sim 1/Ns$, rather than the nominal frequency threshold f_0.

To streamline notation, it will be useful to summarize this behavior by defining an effective fitness cost,

$$s^*(f_0) \sim \begin{cases} \frac{1}{Nf_0} & \text{if } Nsf_0 \ll 1 \\ s & \text{if } Nsf_0 \gg 1 \end{cases}\,, \quad (23)$$

such that Eq. (22) can be written in the compact form

$$p(f|s^*) \sim \frac{N\mu\cdot\theta\left(\frac{1}{Ns^*}-f\right)}{f}\,, \quad (24)$$

or alternatively,

$$\Pr[f\sim 1/Ns^*] \sim N\mu\,, \quad (25)$$

where the dependence on the underlying fitness cost s is completely encapsulated in the definition of s^*. This notation emphasizes that the present day frequencies of neutral and deleterious mutations will be essentially identical to each other, given an appropriate choice of s^*.

However, this correspondence between neutral and deleterious mutations starts to break down when we consider the historical trajectories of these mutations backward in time. The key difference is that natural selection will prevent deleterious mutations from growing much larger than $f_{\rm sel}{\sim}1/Ns$ at *any* point during their lifetime, and not only at the point of observation (Fig. 1). This distinction has a negligible impact on the upward trajectories that dominate $p(f|f_0)$, since their maximum sizes are by definition bounded by the present day frequency f. However, the historical action of natural selection has a much stronger impact on the downward trajectories in Fig. 1, since it limits their peak frequencies to a maximum size of order $f_{\max}{\sim}1/Ns$. This frequency threshold implies a corresponding bound on the maximum age of a deleterious mutation of order $T_{\max}{\sim}1/s$, which alters the scaling of quantities like the average age of a rare mutation,

$$\langle T|f\rangle \sim \begin{cases} Nf\log(1/f) & \text{if } f \ll 1 \ll \frac{1}{Ns}\,, \\ Nf\log(1/Nsf) & \text{if } f \ll \frac{1}{Ns} \ll 1, \\ 1/s & \text{if } f \sim \frac{1}{Ns} \ll 1. \end{cases} \tag{26}$$

These differences will play an important role when we consider the dynamics of linkage disequilibrium below.

Two-locus dynamics

We are now in a position to analyze the joint behavior of a pair of genetic loci, which will allow us to develop a similar heuristic picture for the dynamics of linkage disequilibrium. To do so, it will be useful to first consider the dynamics in the absence of recombination, when mutations at the two loci evolve in a completely asexual manner. Linkage disequilibrium is defined when both sites harbor segregating mutations at the same time ($f_A, f_B > 0$). In the weak mutation limit ($N\mu \ll 1$), these variants must trace back to a single pair of mutation events at each of the two sites. There are only two different ways that these mutations can occur:

Separate mutations. In the simplest scenario, a mutation will first occur in the wildtype background at one of the two sites, and then a second mutation will arise in a different wildtype background while the first mutation is still segregating in the population (Fig. 2A). We refer to this scenario as the *separate mutations* case, since it involves only single-mutant haplotypes ($f_A = f_{Ab}$, $f_B = f_{aB}$), while the double mutant is absent ($f_{AB} = 0$). At low mutation frequencies ($f_A, f_B \ll 1$), these single-mutant haplotypes will be approximately independent of each other, and can therefore be predicted from the single-locus dynamics described above. Equation (25) then shows that with probability

$$\Pr[\text{separate}]{\sim}(N\mu)^2\,, \tag{27a}$$

the four haplotype frequencies will reach typical sizes of order

$$\begin{pmatrix} f_{ab} & f_{aB} \\ f_{Ab} & f_{AB} \end{pmatrix}_{\text{separate}} \sim \begin{pmatrix} 1 & \frac{1}{Ns_B^*} \\ \frac{1}{Ns_A^*} & 0 \end{pmatrix}\,, \tag{27b}$$

where s_A^* and s_B^* are defined as in Eq. (23) above. This coarse-grained sampling distribution allows us to quickly estimate the contribution to various moments, e.g.

$$\langle (f_{AB} - f_A f_B)^k \rangle_{\text{separate}} \sim \frac{(-1)^k (N\mu)^2}{(Ns_A^*)^k\,(Ns_B^*)^k}\,. \tag{28}$$

Nested mutations. In the alternative scenario, a mutation at the second locus can be produced by the original mutant lineage rather than the wildtype background (Fig. 2B,C). We refer to this scenario as the *nested mutations* case, since it produces double mutant haplotypes ($f_A = f_{Ab} + f_{AB}$, $f_B = f_{AB}$) without any single mutants at the other genetic locus ($f_{aB} = 0$). At low mutation frequencies ($f_A, f_B \ll 1$), we expect that these nested mutations will occur far less frequently than separate mutations, since the mutant lineage will produce mutations at a much lower rate ${\sim}N\mu f_{Ab} \ll N\mu$. To understand these contributions, it is necessary to integrate over the historical frequencies of the Ab haplotype (f_{Ab}^0) at the various historical times T that the second mutation could have occurred. We can write this integral in the general form,

$$\begin{aligned} p(f_{Ab}, f_{AB})\,df_{Ab}\,df_{AB} \sim \iint p(f_{Ab}^0)df_{Ab}^0 \cdot N\mu f_{Ab}^0 dT \\ \times p(f_{Ab}^0 \to f_{Ab}|T)df_{Ab} \cdot p\left(\frac{1}{N} \to f_{AB}|T\right) df_{AB}\,, \end{aligned} \tag{29}$$

where $p(f_{Ab}^0)$ denotes the equilibrium frequency distribution at the first site, and $p(f_i^0 \to f_i|T)$ denotes the probability density of transitioning from the historical frequency f_i^0 to the present-day frequency f_i in time T. Note that since f_{Ab}^0 represents the *historical* frequency of the Ab haplotype, it is not *directly* constrained by the present-day frequency threshold, $f_{Ab}, f_{AB} \lesssim f_0$, but is instead constrained *indirectly* through the dynamics in the $p(f_{Ab}^0 \to f_{Ab}|T)$ term. This distinction will become important below.

To gain a more intuitive understanding of Eq. (29), it will be useful to further distinguish between the contributions of relatively recent nested mutations ($T \sim 1/s_{AB}^*$) and those that arose in the distant past ($T \gg 1/s_{AB}^*$), where s_{AB}^* is the double mutant analogue of s_A^* and s_B^* in Eq. (27b). For simplicity, we will restrict our attention to regimes where $s_{AB}^* \gtrsim s_A^*, s_B^*$, such that the maximum lifetimes of the single mutants (${\sim}1/s_i^*$) are at least as long as the double mutants. This regime includes the traditional additive case ($\epsilon = 0$), as well as cases with strong synergistic ($\epsilon > 0$) and moderately antagonistic ($\epsilon < 0$) epistasis, and will therefore capture much of the interesting behavior.

Recent nested mutations. When the age of the double mutant is of order $T{\sim}1/s_{AB}^*$, the characteristic dynamics will be dominated by upward trajectories, in which the double mutants reach their maximum frequency ${\sim}1/Ns_{AB}^*$ near the point of observation (Fig. 2B). Since $T{\lesssim}1/s_A^*$, the historical frequency of the Ab haplotype cannot be much larger than $f_{Ab}^0{\sim}1/Ns_A^*$, since it would otherwise be unlikely to drift back down to this threshold by the time of observation. This suggests that recent nested mutations will occur with a total probabil-

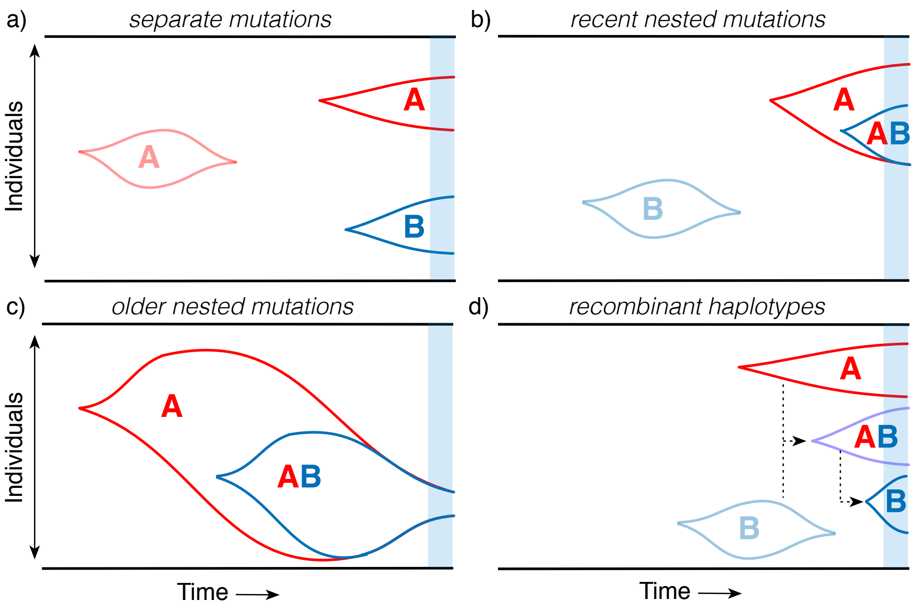

FIG. 2 **Schematic of different lineage dynamics that contribute to linkage disequilibrium.** (a) Separate mutations: A and B mutations arise on independent wildtype backgrounds and are both still segregating at the time of observation (blue region). (b) Recent nested mutations: a double mutant (AB) is produced by a single-mutant background (A) in the recent past, and both haplotypes are still segregating at the time of observation. (c) Older nested mutations: a double mutant is produced by a larger single-mutant lineage in the distant past, but drifts back down to lower frequencies by the time of observation. (d) Recombination produces double mutant lineages from single mutant lineages, and vice versa.

ity

$$
\begin{aligned}
\Pr\left[\begin{matrix}\text{recent}\\ \text{nested}\end{matrix}\right] &\sim \iint_{\substack{T\sim\frac{1}{s^*_{AB}}\\ f^0_{Ab}\lesssim\frac{1}{Ns^*_A}}} \frac{N\mu\, df^0_{Ab}}{f^0_{Ab}}\cdot N\mu f^0_{Ab} dT\cdot\frac{1}{T}\,,\\
&\sim\frac{(N\mu)^2}{Ns^*_A}\,, \qquad\qquad (30\text{a})
\end{aligned}
$$

and that the typical haplotype frequencies will be of order

$$
\begin{pmatrix} f_{ab} & f_{aB}\\ f_{Ab} & f_{AB}\end{pmatrix}_{\substack{\text{recent}\\ \text{nested}}} \sim \begin{pmatrix} 1 & 0\\ \frac{1}{Ns^*_A} & \frac{1}{Ns^*_{AB}}\end{pmatrix}\,. \qquad (30\text{b})
$$

An analogous distribution exists for recent nested mutations that arise on an aB background. As expected, the total probability of these events is much smaller than the separate mutations case. However, the smaller probability is balanced by the larger typical values of $D=f_{AB}-f_Af_B$ that occur in this case, such that the total contribution to the corresponding moments,

$$
\langle (f_{AB}-f_Af_B)^k\rangle_{\substack{\text{recent}\\ \text{nested}}} \sim \frac{(Ns^*_A+Ns^*_B)(N\mu)^2}{(Ns^*_A\cdot Ns^*_B)\,(Ns^*_{AB})^k}\,, \qquad (31)
$$

is often of equal or larger magnitude than the separate mutations case in Eq. (28) above.

Older nested mutations. In contrast, the characteristic dynamics of older mutations ($T\gg 1/s^*_{AB}$) will be dominated by downward trajectories that previously reached much higher frequencies in the distant past (Fig. 2C). We can analyze these trajectories using the same techniques that we used to study the average ages of neutral and deleterious mutations above. To streamline our notation, it will be useful to introduce a timescale for the maximum possible age of a double mutant,

$$
T^{AB}_{\text{max}}\sim\begin{cases} 1/s_{AB} & \text{if } Ns_{AB}\gg 1,\\ N & \text{if } Ns_{AB}\ll 1,\end{cases} \qquad (32)
$$

which corresponds to a maximum historical size of order $f_{\text{max}}{\sim}T^{AB}_{\text{max}}/N$. This timescale depends on the true fitness cost s_{AB} (rather than s^*_{AB}) because the historical frequencies are not directly constrained by the present-day frequency threshold. When $Ns_{AB}f_0\gg 1$, this maximum age is located in the recent past ($T^{AB}_{\text{max}}{\sim}1/s^*_{AB}$), which implies that there is a negligible chance of producing older nested mutations. In contrast, when $Ns_{AB}f_0\ll 1$, there will be a large range of timescales ($1/s^*_{AB}\lesssim T\lesssim T^{AB}_{\text{max}}$) where older nested mutations can arise.

These older double mutants will have a much smaller chance of surviving to the present while also maintaining a present-day frequency less than $\sim 1/Ns^*_{AB}$:

$$
\Pr\left[0<f_{AB}\lesssim\tfrac{1}{Ns^*_{AB}}\middle|\,T\right]\sim\frac{1}{T}\cdot\frac{\frac{1}{Ns^*_{AB}}}{\frac{T}{N}}\sim\frac{1}{T^2s^*_{AB}}\,. \qquad (33)
$$

Similar constraints apply to the single mutant trajectories that generate these double mutant haplotypes. If the historical frequency f_{Ab}^0 is much larger than $\sim T/N$, the mutation is unlikely to drift back down to its present-day frequency threshold ($f_{Ab} \sim 1/Ns_A^*$) by the time of observation. On the other hand, historical frequencies much smaller than $\sim T/N$ are highly likely to go extinct before the present day (Fig. 2C), and will automatically satisfy the frequency threshold at the time of sampling. This suggests that historical frequencies $f_{Ab}^0 \lesssim T/N$ will provide the dominant contribution to the total probability of this scenario. Combining these observations, we conclude that older nested mutations occur with a total probability

$$\begin{aligned}\Pr\left[\begin{matrix}\text{older}\\\text{nested}\end{matrix}\right] &\sim \iint_{\substack{f_{Ab}^0 \lesssim \frac{T}{N}\\ T \lesssim T_{\max}^{AB}\\ T \gtrsim \frac{1}{s_{AB}^*}}} \frac{N\mu\, df_{Ab}^0}{f_{Ab}^0} \cdot N\mu f_{Ab}^0 dT \cdot \frac{1}{s_{AB}^* T^2} \\ &\sim \frac{(N\mu)^2}{Ns_{AB}^*} \log\left(T_{\max}^{AB} s_{AB}^*\right) \,, \qquad (34a)\end{aligned}$$

and that the typical haplotype frequencies will be of order

$$\begin{pmatrix} f_{ab} & f_{aB} \\ f_{Ab} & f_{AB} \end{pmatrix}_{\substack{\text{older}\\\text{nested}}} \sim \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{Ns_{AB}^*} \end{pmatrix} . \qquad (34b)$$

Note that the total probability of these events is still much smaller than the probability of arising on separate genetic backgrounds. However, the logarithmic factor ensures that this probability will be larger than the contribution from recent nested mutations whenever $Ns_{AB}f_0 \ll 1$.

Estimating linkage disequilibrium. We now have all the ingredients necessary to calculate frequency-resolved LD statistics like $\sigma_d^k(f_0)$ in Eq. (7). For the denominator, we note that the magnitude of $f_A(1-f_A)f_B(1-f_B)$ is roughly the same regardless of whether the mutations occur on nested or separate backgrounds. However, since the separate backgrounds case is much more likely, this scenario will provide the dominant contribution to the average in the denominator, so that

$$\langle f_A(1-f_A)f_B(1-f_B)\rangle \sim \frac{1}{Ns_A^*} \cdot \frac{1}{Ns_B^*} \cdot (N\mu)^2 \,. \qquad (35)$$

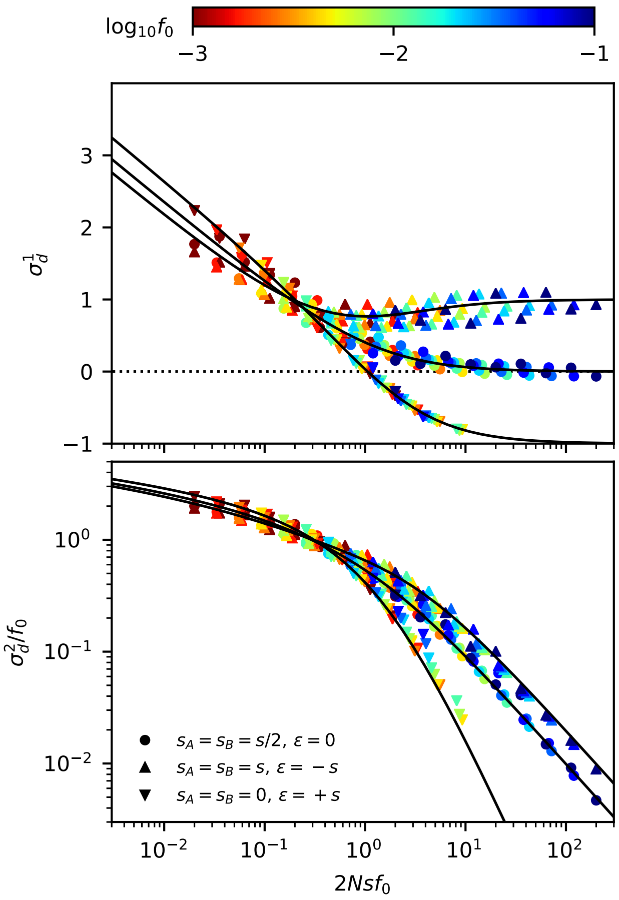

FIG. 3 **Frequency-resolved LD between deleterious mutations as a function of the scaled fitness cost of the double mutant.** Top: The signed LD moment, $\sigma_d^1(f_0)$, in Eq. (7) is depicted for pairs of nonrecombining loci with additive ($\epsilon = 0$), antagonistic ($\epsilon < 0$) and synergistic ($\epsilon > 0$) epistasis, which were chosen to have the same total cost for the double mutant ($s_{AB} = s$). Symbols denote the results of forward-time simulations (Appendix A) across a range of parameters with $s > 10/N$, and each symbol is colored by the corresponding value of f_0. The solid lines shows the theoretical prediction from Eq. (C8). Bottom: An analogous figure for the squared LD moment, $\sigma_d^2(f_0)$, where solid lines show the theoretical predictions from Eq. (C10). The "data collapse" in both panels indicates that frequency-weighted LD is primarily determined by the compound parameters Nsf_0 and $N\epsilon f_0$. Weak scaled fitness costs ($Nsf_0 \lesssim 1$) lead to an excess of coupling linkage ($\sigma_d^1 > 0$), which qualitatively resembles the effects of antagonistic epistasis ($\epsilon < 0$).

In contrast, the numerator of $\sigma_d^2(f_0)$ will usually be dominated by the contributions from nested mutations, since the double mutant frequency enters as a lower power in the linkage disequilibrium coefficient $D = f_{AB} - f_A f_B$. The precise form of this contribution will depend on the typical ages of the nested mutations, as described by Eqs. (27) and (30) above. By combining these results with the denominator term in Eq. (35), we find that

$$\sigma_d^2(f_0) \sim \begin{cases} f_0 \log\left(\frac{1}{f_0}\right) & \text{if } Ns_{AB} \ll 1, \\ f_0 \log\left(\frac{1}{Ns_{AB}f_0}\right) & \text{if } Ns_{AB}f_0 \ll 1, \\ \frac{1 - \frac{\epsilon}{s_{AB}} + \frac{1}{Ns_{AB}f_0}}{Ns_{AB}} & \text{if } Ns_{AB}f_0 \gg 1, \end{cases} \qquad (36)$$

which strongly depends on the magnitude of the scaled selection strength $Ns_{AB}f_0$. In the absence of epistasis ($\epsilon = 0$), Eq. (36) shows that strongly deleterious mutations ($s_A{\approx}s_B{\approx}s{\gg}1/N$) will generally have lower LD than neutral mutations with comparable present-day frequencies ($f_{0,\mathrm{eff}}{\sim}1/Ns$) (Fig. 3). Equation (36) shows that the direction of this effect is qualitatively similar to the effects of synergistic epistasis ($\epsilon > 0$). These differences between neutral and deleterious mutations are qualitatively different from the behavior observed for the site frequency spectrum in Eq. (24), where fitness costs could be absorbed by a simple present-day frequency threshold (Fig. 1). Our lineage-based picture shows that these differences in LD primarily reflect the contributions of older nested mutations (Fig. 2C), which are sensitive to the effects of natural selection long before the present day.

An analogous argument can be used to calculate the first moment $\sigma_d^1(f_0)$. We find that

$$\sigma_d^1(f_0) \sim \begin{cases} \log\left(\frac{1}{f_0}\right) & \text{if } Ns_{AB} \ll 1, \\ \log\left(\frac{1}{Ns_{AB}f_0}\right) & \text{if } Ns_{AB}f_0 \ll 1, \\ -\frac{\epsilon}{s_{AB}} & \text{if } Ns_{AB}f_0 \gg 1, \end{cases} \tag{37}$$

which closely parallels the three regimes that emerge for $\sigma_d^2(f_0)$ in Eq. (36). Once again, we observe a dramatic difference between neutral and deleterious mutations that cannot be captured by an effective frequency threshold $f_{0,\mathrm{eff}}{\sim}1/Ns$. In this case, we see that the logarithmic behavior in the neutral limit is associated with an excess of coupling linkage ($f_{AB} > f_A f_B$), which qualitatively resembles the effects of antagonistic epistasis ($\epsilon < 0$). These results emphasize the importance of older nested mutations in shaping contemporary patterns of LD among tightly linked loci.

Incorporating recombination. We can now ask how small amounts of recombination start to alter the basic picture described above. Recombination cannot change the rates at which separate or nested mutations are initially produced, but it can have a dramatic impact on the subsequent haplotype dynamics after these mutations arise.

For example, in the nested mutations case, recombination will start to break up the AB haplotype at a per capita rate R, creating recombinant Ab and aB offspring (Fig. 2D). From the perspective of the f_{AB} lineage, this loss of individuals through recombination will resemble an effective fitness cost, which can be absorbed in an effective selection coefficient for the double mutant,

$$s_{\mathrm{AB,r}} \equiv s_A + s_B + \epsilon + R\,. \tag{38}$$

When $R \ll s_{AB}$, this loss of individuals through recombination will have a negligible impact on f_{AB}, but it will become the primary limiting factor on the lineage size when $R \gg s_{AB}$. For sufficiently low rates of recombination (evaluated self-consistently below), the recombinant offspring of AB lineages are unlikely reach high enough frequencies to influence the linkage disequilibrium coefficient $D = f_{AB} - f_A f_B$. This suggests that we can approximate the future dynamics of the f_{AB} lineage using the results for the asexual case above, but with Eq. (38) replacing s_{AB}.

Similarly, in the separate mutations scenario, recombination events between Ab and aB haplotypes will create recombinant AB haplotypes at a total rate $NRf_{Ab}f_{aB}$ per generation. In this case, the loss of individuals due to recombination is significantly smaller than for the AB lineages above, since the per capita rates of recombination for the single-mutant lineages are suppressed by additional factors of f_{aB} or f_{Ab}. However, these rare recombination events can still have a large effect on linkage disequilibrium if they happen to seed a lucky AB lineage that drifts to observable frequencies. We can calculate the total probability of these events using a generalization of the approach that we used for nested mutations above.

In this recombinant scenario, the dominant contributions will come from relatively recent recombination events ($T{\sim}1/s^*_{\mathrm{AB,r}}$), which reach their maximum typical size ($f_{AB}{\sim}1/s^*_{\mathrm{AB,r}}$) near the time of observation (Fig. 2D). As above, the historical frequency of the Ab and aB haplotypes cannot be much larger than $f^0_{Ab}{\sim}1/Ns^*_A$ and $f^0_{aB}{\sim}1/Ns^*_B$ at this timepoint, otherwise they would be unlikely to drift back down to these thresholds by the time of observation. This suggests that recombinant double mutant will occur with a total probability

$$\begin{aligned} \Pr\begin{bmatrix}\text{recomb.}\\ \text{double}\end{bmatrix} &\sim \iiint_{\substack{T\sim\frac{1}{s^*_{AB}} \\ f^0_{Ab}\lesssim\frac{1}{Ns^*_A} \\ f^0_{aB}\lesssim\frac{1}{Ns^*_B}}} \left[\frac{N\mu\, df^0_{Ab}}{f^0_{Ab}}\cdot\frac{N\mu\, df^0_{aB}}{f^0_{aB}}\right. \\ &\qquad\qquad \left.\times NRf^0_{Ab}f^0_{aB}\, dT\cdot\frac{1}{T}\right]\,, \\ &\sim \frac{NR(N\mu)^2}{Ns^*_A\cdot Ns^*_B}\,, \end{aligned} \tag{39a}$$

and that the haplotype frequencies will be of order

$$\begin{pmatrix} f_{ab} & f_{aB} \\ f_{Ab} & f_{AB}\end{pmatrix}_{\substack{\text{recomb.}\\ \text{double}}} \sim \begin{pmatrix} 1 & \frac{1}{Ns^*_A} \\ \frac{1}{Ns^*_B} & \frac{1}{Ns^*_{\mathrm{AB,r}}}\end{pmatrix}\,. \tag{39b}$$

Similar to the nested mutations case above, the total probability of the recombinant scenario is much smaller than the probability of the separate mutations case. However, as long as $f_{AB}{\sim}1/Ns^*_{\mathrm{AB,r}}$ is much larger than $f_{Ab}f_{aB}{\sim}1/(Ns^*_A\cdot Ns^*_B)$ — which will be true for all but the highest recombination rates — the smaller probability of this scenario will be counterbalanced by its significantly larger values of $D = f_{AB} - f_A f_B$. By combining these results with our previous formulae for nested and separate mutations above, we can obtain an analogous set of predictions for the frequency-resolved LD statistics,

$$\sigma_d^2(f_0) \sim \begin{cases} f_0 \log\left(\frac{1}{f_0}\right) & \text{if } Ns_{\mathrm{AB,r}} \ll 1, \\ f_0 \log\left(\frac{1}{Ns_{\mathrm{AB,r}} f_0}\right) & \text{if } Ns_{\mathrm{AB,r}} f_0 \ll 1, \\ \frac{1 - \frac{\epsilon}{s_{\mathrm{AB,r}}} + \frac{1}{Ns_{\mathrm{AB,r}} f_0}}{Ns_{\mathrm{AB,r}}} & \text{if } Ns_{\mathrm{AB,r}} f_0 \gg 1, \end{cases} \tag{40a}$$

and

$$\sigma_d^1(f_0) \sim \begin{cases} \log\left(\frac{1}{f_0}\right) & \text{if } Ns_{\mathrm{AB,r}} \ll 1, \\ \log\left(\frac{1}{Ns_{\mathrm{AB,r}} f_0}\right) & \text{if } Ns_{\mathrm{AB,r}} \ll 1, \\ -\frac{\epsilon}{s_{\mathrm{AB,r}}} & \text{if } Ns_{\mathrm{AB,r}} \gg 1, \end{cases} \tag{40b}$$

which are valid in the presence of recombination.

For the special case neutral mutations ($s_A = s_B = \epsilon = 0$), these results take on a particularly simple form:

$$\sigma_d^2(f_0) \sim \begin{cases} f_0 \log\left(\frac{1}{f_0}\right) & \text{if } NR \ll 1, \\ f_0 \log\left(\frac{1}{NRf_0}\right) & \text{if } NRf_0 \ll 1, \\ \frac{1}{NR} & \text{if } NRf_0 \gg 1, \end{cases} \tag{41a}$$

and

$$\sigma_d^1(f_0) \sim \begin{cases} \log\left(\frac{1}{f_0}\right) & \text{if } NR \ll 1, \\ \log\left(\frac{1}{NRf_0}\right) & \text{if } NRf_0 \ll 1, \\ 0 & \text{if } NRf_0 \gg 1. \end{cases} \tag{41b}$$

These expressions constitute frequency-resolved analogues of the LD decay curves that are used to estimate recombination rates in genomic data (Fig. 4). In this case, we see that the behavior of the LD curves is strongly dependent on the compound parameter NRf_0. When $NRf_0 \gg 1$, we recover the well-known $\sigma_d^2 \sim 1/NR$ scaling of Eq. (2), in which neither coupling or repulsion linkage is favored ($\sigma_d^1 \approx 0$) (Ohta and Kimura, 1971; Song and Song, 2007). However, when $NRf_0 \ll 1$, $\sigma_d^2(f_0)$ no longer saturates at a constant value, as in Eq. (2), but instead displays a new logarithmic dependence similar to asexual case above. These quasi-asexual dynamics are accompanied by high levels of coupling linkage ($\sigma_d^1 \gtrsim 1$), reflecting the important contributions of older nested mutations. Interestingly, our results show that the transition to this mutation-dominated regime can occur even for nominally high rates of recombination ($NR \gg 1$), provided that the frequency scale f_0 is chosen to be sufficiently small ($NRf_0 \ll 1$). This highlights the utility of frequency-resolved LD statistics for probing the underlying timescales of recombination process — a topic that we will explore in more detail below.

Transition to Quasi-Linkage Equilibrium (QLE). The results above assumed that double mutants provide the dominant contribution to $D = f_{AB} - f_A f_B$ when they reach their maximum typical frequencies. This will be a good approximation in the recombinant scenario provided that

$$\frac{1}{Ns^*_{\mathrm{AB,r}}} \gg \frac{1}{Ns^*_A} \cdot \frac{1}{Ns^*_B}\,, \tag{42}$$

which reduces to the simpler condition

$$NRf_0^2 \ll 1\,, \tag{43}$$

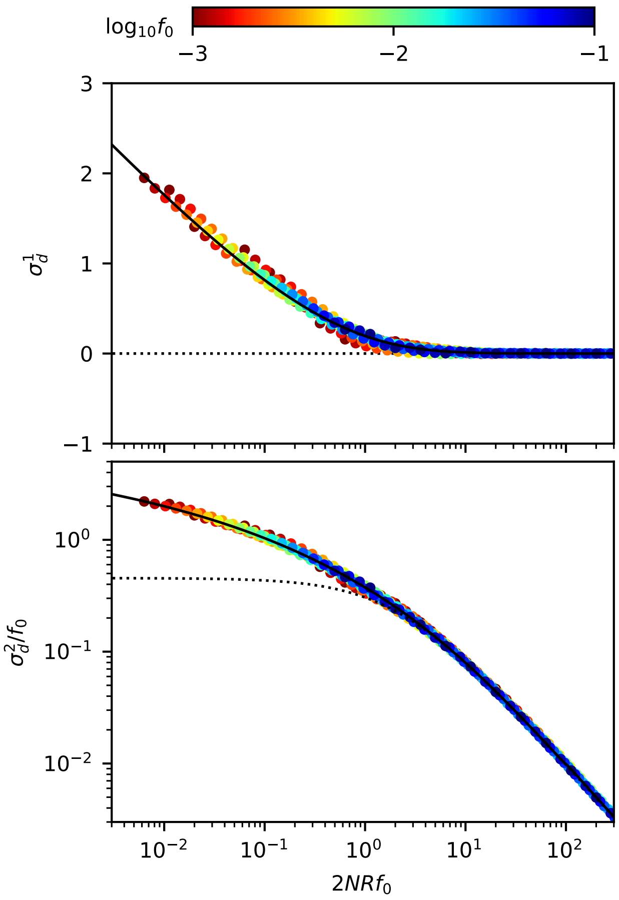

FIG. 4 **Frequency-resolved LD between neutral mutations as a function of the scaled recombination rate.** An analogous version of Fig. 3, showing the first (top) and second (bottom) LD moments in Eq. (7) for pairs of neutral mutations with a range of recombination rates, $R > 2/N$. As above, symbols denote the results of forward time simulations, and solid lines denote the theoretical predictions from Eqs. D23 (top) and D24 (bottom). Dashed lines show the classical predictions for the $f_0 \to \infty$ limit (Ohta and Kimura, 1971). The "data collapse" in both panels indicates that frequency-weighted LD is primarily determined by compound parameter NRf_0. Low scaled recombination rates ($NRf_0 \lesssim 1$) lead to an excess of coupling linkage ($\sigma_d^1 > 0$), which qualitatively resembles the effects of antagonistic epistasis ($\epsilon < 0$).

for a pair of neutral mutations. At low frequencies ($f_0 \ll 1$), this condition will generally be satisfied even for large recombination rates ($NRf_0 \gg 1$), which are located deep into the recombination-dominated regions of the LD curves in Eq. (41). This suggests that these previous expressions will be valid across a broad parameter range, which is sufficient to capture the transition from mutation-dominated to recombination-dominated behavior ($NRf_0 \sim 1$). Nevertheless, for sites that are separated by sufficiently large coordinate distances ℓ, recombination rates may eventually grow large enough that $NRf_0^2 \gtrsim 1$, where our previous analysis starts to break down.

Fortunately, we can obtain a relatively complete picture of this transition by taking advantage of the large gap between NRf_0 and NRf_0^2 that emerges when $f_0 \ll 1$, and restricting our attention to cases where $NRf_0 \gg 1$. In this limit, the typical lifetimes of recombinant double mutants ($T \sim 1/R$) are much shorter than the lifetimes of the single mutant lineages that produce them ($T \sim Nf_0$). This suggests that f_{Ab} and f_{aB} will be effectively "frozen" throughout the lifetime of an individual recombinant lineage, and that the production rate of these lineages will resemble a mutation process with an overall rate $\sim NRf_0^2$. When $NRf_0^2 \ll 1$, recombinant f_{AB} lineages will be produced only rarely, and can occasionally fluctuate to frequencies of order $\sim 1/NR$ before they go extinct. This is sufficient to recover the familiar $\sigma_d^2 \sim 1/NR$ scaling,

$$\sigma_d^2(f_0) \sim \frac{NRf_0^2 \cdot \left(\frac{1}{NR}\right)^2}{f_0^2} \sim \frac{1}{NR}\,, \tag{44}$$

that we observed in Eq. (41).

On the other hand, when $NRf_0^2 \gg 1$, many recombinant lineages will be produced every generation, and the total double mutant frequency f_{AB} will reach sizes much larger than $\sim 1/NR$. In this case, the double mutant frequency will grow to the point where the production rate of new recombinants ($\sim NRf_0^2$) is exactly balanced by their loss due to further recombination with the wild-type population ($\sim NRf_{AB}$). This occurs when

$$f_{AB} \sim f_0^2\,, \tag{45}$$

which is equivalent to the condition that the A and B mutations are in linkage equilibrium ($f_{AB} = f_A f_B$). In this way, we see that the $NRf_0^2 \gg 1$ regime can be identified with the traditional Quasi-Linkage Equilibrium (QLE) regime of multilocus population genetics, in which the haplotype frequencies remain close to the typical values expected under linkage equilibrium. Genetic drift will still drive fluctuations around the average value in Eq. (45), with a magnitude that is inversely proportional to the typical number of recombinant lineages that contribute to the total frequency:

$$\left(\frac{\delta f_{AB}}{f_{AB}}\right)^2 \sim \left(\frac{f_{AB}}{\frac{1}{NR}}\right)^{-1} \sim \frac{1}{NRf_0^2}\,. \tag{46}$$

Interestingly, this behavior leads to identical predictions for the two lowest order LD statistics, $\sigma_d^1(f_0)$ and $\sigma_d^2(f_0)$, that we obtained in the $f_0^{-1} \ll NR \ll f_0^{-2}$ limit above. However, as we will demonstrate below, the differences between these two regimes will become apparent when considering higher moments (or other properties of the joint haplotype distribution), due to the dramatic differences in the typical fluctuations of f_{AB}. In this way, we see that low frequency mutations give rise to an entirely new regime of behavior ($f_0^{-1} \ll NR \ll f_0^{-2}$), in which previous notions of quasi-asexuality or quasi-linkage equilibrium do not apply. We will refer to this regime as the *clonal recombinant regime*, which reflects the fact that double mutants are primarily caused by rare recombinant lineages that drift to observable frequencies one-by-one. We will explore the consequences of these dynamics in more detail below.

FORMAL ANALYSIS

We now turn to a formal derivation of the heuristic results presented above. To implement our conditioning scheme for the maximum frequencies of the two mutations ($f_A, f_B \lesssim f_0$), we will focus on the joint generating function of the unconditioned process,

$$H(x,y,z,t) = \left\langle e^{-\frac{xf_{Ab}(t)}{f_0} - \frac{yf_{aB}(t)}{f_0} - \frac{zf_{AB}(t)}{f_0}} \right\rangle\,, \tag{47}$$

such that the weighted moments follow from the identity

$$\left\langle f_{Ab}^i f_{aB}^j f_{AB}^k e^{-\frac{f_A+f_B}{f_0}} \right\rangle = f_0^p \left.\frac{(-1)^p \partial^p H}{\partial x^i \partial y^j \partial z^k}\right|_{\substack{x=1\\y=1\\z=2\\t=\infty}}\,, \tag{48}$$

with $p = i+j+k$. We can also define an *effective conditional distribution* using a similar weighting scheme,

$$\begin{aligned} H(x,y,z|f_0) &\equiv \frac{\left\langle e^{-\frac{xf_{Ab}+yf_{aB}+zf_{AB}}{f_0}} \cdot e^{-\frac{f_A+f_B}{f_0}} \right\rangle}{\left\langle e^{-\frac{f_A+f_B}{f_0}} \right\rangle}\,, \\ &= \lim_{t\to\infty} \frac{H(x+1,y+1,z+2,t)}{H(1,1,2,t)}\,, \end{aligned} \tag{49}$$

such that the weighted moments can also be obtained directly from derivatives of $H(x,y,z|f_0)$. Thus, for this special choice of weighting function, the conditional moments can be straightforwardly calculated from solutions of the unconditioned generating function, $H(x,y,z,t)$. By differentiating Eq. (47) with respect to time and applying the stochastic dynamics in Eq. (8), we find that

$H(x,y,z,t)$ satisfies the partial differential equation,

$$\begin{aligned}\frac{\partial H}{\partial \tau} &= -\left(\gamma_A x + x^2\right)\frac{\partial H}{\partial x} - \left(\gamma_B y + y^2\right)\frac{\partial H}{\partial y} \\ &\quad - \left[\gamma_{AB} z + z^2 - \rho(x{+}y)\right]\frac{\partial H}{\partial z} - \theta(x{+}y)H \\ &\quad + \theta f_0 z\left(\frac{\partial H}{\partial x} + \frac{\partial H}{\partial y}\right) - \rho f_0 (z{-}x{-}y)\frac{\partial^2 H}{\partial x \partial y}\,,\end{aligned} \tag{50}$$

with the initial condition $H(x,y,z,0) = 1$, where we have defined a collection of scaled variables,

$$\begin{gathered}\theta = 2N\mu\,, \quad \tau = t/2Nf_0\,, \quad \rho = 2NRf_0\,, \\ \gamma_A = 2Ns_A f_0\,, \quad \gamma_B = 2Ns_B f_0\,, \quad \gamma_\epsilon = 2N\epsilon f_0\,, \\ \gamma_{AB} = \gamma_A + \gamma_B + \gamma_\epsilon + \rho\,.\end{gathered} \tag{51}$$

Here we have used the conventional symbols θ, ρ, and γ_i to define scaled rates of mutation, recombination, and selection, respectively. Note that in the latter two cases, we have defined these scaled variables to include an additional factor of f_0 in order to match the key control parameters that we obtained in our heuristic analysis above. Motivated by these results, we will also restrict our attention to scenarios where $s, r \gg 1/N$, but where the scaled parameters ρ and γ can be either large or small compared to one. This will ensure that the maximum historical frequencies remain sufficiently small that the branching process approximation remains valid, while still capturing the full range of the qualitative behavior identified above.

Perturbation expansion for small f_0

The partial differential equation in Eq. (50) is difficult to solve in the general case. To make progress, we will focus on a perturbation expansion in the limit that θ and f_0 are both small compared to one, using the series ansatz

$$H(x,y,z,\tau) = 1 + \sum_{i,j=0}^{\infty} \theta^i f_0^j H_{i,j}(x,y,z,\tau)\,, \tag{52}$$

with $H_{i,j}(x,y,z,0) = 0$. The first few terms in this expansion are calculated in Appendix B. The first order contributions are simply a product of the corresponding single-locus distributions,

$$\begin{aligned}H(x,y,z,\tau) &\approx 1 - \theta H_A(x,\tau) \\ &\qquad - \theta H_B(y,\tau) + \mathcal{O}(\theta^2)\,,\end{aligned} \tag{53a}$$

where we have defined

$$\begin{aligned}H_A(x,\tau) &\equiv \log\left[1 + \frac{x(1-e^{-\gamma_A \tau})}{\gamma_A}\right]\,, \\ H_B(y,\tau) &\equiv \log\left[1 + \frac{x(1-e^{-\gamma_B \tau})}{\gamma_B}\right]\,.\end{aligned} \tag{53b}$$

The conditional distribution in Eq. (49) then follows as

$$\begin{aligned}H(x,y,z|f_0) &\approx 1 - \theta\log\left(1 + \frac{x}{\gamma_A + 1}\right) \\ &\qquad - \theta\log\left(1 + \frac{y}{\gamma_B + 1}\right) + \mathcal{O}(\theta^2)\,.\end{aligned} \tag{54}$$

This distribution has a well-defined value even when $\gamma_i = 0$, which shows how the frequency weighting in Eq. (49) can eliminate the well-known divergence of the neutral branching process when $\tau \to \infty$. We also see that the resulting distributions are equivalent to the *unconditioned* frequency spectrum of a deleterious mutation with an effective selection coefficient $\gamma^* = \gamma + 1$, which has a well-known form,

$$p(f|f_0) \approx \frac{2N\mu \cdot e^{-2Nsf - f/f_0}}{f}\,. \tag{55}$$

This constitutes a quantitative version of the heuristic result in Eq. (22), and confirms our previous intuition that the deleterious site frequency spectrum can be mimicked by neutral mutations with an appropriate choice of f_0.

By definition, these first order solutions do not provide any information about linkage disequilibrium between the two loci, which only starts to enter at order θ^2. A more detailed calculation in Appendix B shows that these next order contributions can be written in the form

$$\begin{aligned}H &\approx 1 - \theta(H_A{+}H_B) + \frac{\theta^2}{2}\left(H_A{+}H_B\right)^2 + \theta^2 f_0 \Upsilon \\ &\quad - \theta^2 f_0 \int_0^\tau d\tau'\,\psi(x,y,z,\tau')\left[\Phi_x(\tau,\tau')\right. \\ &\qquad\qquad \left.+\Phi_y(\tau,\tau') + \rho\Phi_x(\tau,\tau')\Phi_y(\tau,\tau')\right]\,,\end{aligned} \tag{56a}$$

where $\Upsilon(x,y,\tau)$ is a function that is independent of z, $\psi(x,y,z,\tau')$ is a solution to the characteristic curve,

$$\begin{aligned}\frac{\partial \psi}{\partial \tau'} &= -\gamma_{AB}\psi - \psi^2 + \frac{\rho\gamma_A x e^{-\gamma_A \tau'}}{\gamma_A + x\left(1 - e^{-\gamma_A \tau'}\right)} \\ &\qquad + \frac{\rho\gamma_B y e^{-\gamma_B \tau'}}{\gamma_B + y\left(1 - e^{-\gamma_B \tau'}\right)}\,,\end{aligned} \tag{56b}$$

with the initial condition $\psi(x,y,z,0) = z$, and $\Phi_x(\tau,\tau')$ and $\Phi_y(\tau,\tau')$ are defined by

$$\begin{aligned}\Phi_x(\tau,\tau') &\equiv \frac{[1 - e^{-\gamma_A(\tau - \tau')}][\gamma_A + x(1 - e^{-\gamma_A \tau'})]}{\gamma_A\left[\gamma_A + x(1 - e^{-\gamma_A \tau})\right]}\,, \\ \Phi_y(\tau,\tau') &\equiv \frac{[1 - e^{-\gamma_B(\tau - \tau')}][\gamma_B + y(1 - e^{-\gamma_B \tau'})]}{\gamma_A\left[\gamma_B + y(1 - e^{-\gamma_B \tau})\right]}\,.\end{aligned} \tag{57}$$

Using this formal solution, the weighted moments then follow as

$$\langle f_A f_B e^{-\frac{f_A + f_B}{f_0}}\rangle \approx \frac{\theta^2 f_0^2}{(\gamma_A + 1)(\gamma_B + 1)} + \mathcal{O}(f_0^3)\,, \tag{58}$$

and

$$\left\langle f_{AB}^{k} e^{-\frac{f_A+f_B}{f_0}} \right\rangle \approx \theta^2 f_0^{k+1} \int_0^\infty d\tau' \, \psi_k(\tau') \left[\Phi_A(\tau') \right. \\ \left. + \Phi_B(\tau') + \rho \Phi_A(\tau') \Phi_B(\tau')\right] , \tag{59a}$$

where we have defined the functions

$$\begin{aligned} \Phi_A(\tau') &\equiv \Phi_x(\tau,\tau')\big|_{\substack{x=1 \\ \tau=\infty}} = \frac{\gamma_A + 1 - e^{-\gamma_A \tau'}}{\gamma_A(\gamma_A+1)} , \\ \Phi_B(\tau') &\equiv \Phi_y(\tau,\tau')\big|_{\substack{y=1 \\ \tau=\infty}} = \frac{\gamma_B + 1 - e^{-\gamma_B \tau'}}{\gamma_B(\gamma_B+1)} , \end{aligned} \tag{59b}$$

and

$$\psi_k(\tau') \equiv (-1)^{k+1} \left. \frac{\partial^k \psi(x,y,z,\tau')}{\partial z^k} \right|_{\substack{x=1 \\ y=1 \\ z=2}} . \tag{59c}$$

Substituting these moments into the definition of $\sigma_d^k(f_0)$, we see that the leading order solution collapses onto a lower dimensional manifold,

$$\frac{\sigma_d^k}{f_0^{k-1}} \approx \Sigma_d^k(\gamma_A, \gamma_B, \gamma_\epsilon, \rho) , \tag{60a}$$

where $\Sigma_d^k(\cdot)$ is a dimensionless function that depends only on the scaled parameters γ_i and ρ:

$$\Sigma_d^k \approx -\delta_{k,1} + (1+\gamma_A)(1+\gamma_B) \int_0^\infty \psi_k(\tau') \left[\Phi_A(\tau') \right. \\ \left. + \Phi_B(\tau') + \rho \Phi_A(\tau') \Phi_B(\tau')\right] d\tau' . \tag{60b}$$

This solution is independent of the mutation rate, as expected, and depends on the frequency scale f_0 only implicitly through the definitions of Σ_d, γ_i, and ρ. This is already an important constraint, as it implies that scenarios with different underlying values of f_0, s_i, and r, but similar values of the scaled parameters γ_i and ρ, must necessarily have similar values of Σ_d^k. Closed form expressions for $\Sigma_d^k(\cdot)$ are more difficult to obtain in the general case, due to the difficulty in solving the differential equation for $\psi(x,y,z,\tau')$ for arbitrary parameter combinations. However, further analytical progress can still be made by examining the behavior of this equation in certain limits.

Non-recombining loci. The simplest behavior occurs in the absence of recombination ($\rho = 0$), when the characteristic curve has an exact solution,

$$\psi(z,\tau') = \frac{\gamma_{AB} z e^{-\gamma_{AB}\tau'}}{\gamma_{AB} + z(1 - e^{-\gamma_{AB}\tau'})} . \tag{61}$$

Substituting this expression into Eqs. (59c) and (60), we find that $\Sigma_d^2(\gamma_A, \gamma_B, \gamma_\epsilon)$ can be expressed as a definite integral,

$$\Sigma_d^2 = \int_0^\infty \frac{e^{-\gamma_{AB}\tau'}(1 - e^{-\gamma_{AB}\tau'})}{4\left(\frac{\gamma_{AB}}{2} + 1 - e^{-\gamma_{AB}\tau'}\right)^3} \\ \times \left[(\gamma_B + 1) \cdot \frac{\gamma_A + 1 - e^{-\gamma_A \tau'}}{\gamma_A} \right. \\ \left. + (\gamma_A + 1) \cdot \frac{\gamma_B + 1 - e^{-\gamma_B \tau'}}{\gamma_B} \right] d\tau' , \tag{62}$$

which is straightforward to evaluate numerically. Analogous integral expressions can be derived for the other moments, $\Sigma_d^k(\gamma_A, \gamma_B, \gamma_\epsilon)$, as well as for the full conditional distribution, $H(x,y,z|f_0)$ (see Appendix C). Asymptotic solutions of these integrals for small and large γ_{AB} show that they have same limiting behavior that we identified in our heuristic analysis above, while the numerical solutions accurately capture the quantitative behavior across the full range of intermediate parameter values (Fig. 3).

Neutral loci. Another important limit occurs for neutral mutations ($\gamma_A = \gamma_B = \gamma_\epsilon = 0$), where the manifold in Eq. (60) reduces to a single parameter curve, $\Sigma_d^k(\rho)$. This is already a useful prediction, since it implies that changes in R can be mimicked by changes in f_0, and vice versa. However, the characteristic curve in Eq. (56b) is now more difficult to solve than in the asexual case, due to the presence of the time-dependent terms, $\rho x/(1+x\tau')$ and $\rho y/(1+y\tau')$. Physically, these terms represent the additional Ab and aB lineages that are created when the AB haplotype recombines with the wildtype background. Fortunately, an exact solution can still be obtained in this case using special functions, which is derived in Appendix D. After substituting this solution into Eq. (59), we find that the scaling function $\Sigma_d^2(\rho)$ can again be expressed as a definite integral,

$$\Sigma_d^2(\rho) = \int_0^\infty \frac{e^{-\zeta}(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)^2} \\ \times \left[(2+\rho) + \frac{(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)} \right] d\zeta , \tag{63}$$

where we have defined the function

$$D(\zeta) \equiv (\rho+\zeta)(2+\rho+\zeta) \left[(1+\rho+\zeta)\rho e^{-\zeta} \right. \\ \left. -1 + \rho e^{\rho}(E_1(\rho) - E_1(\rho+\zeta)) \right] , \tag{64}$$

and $E_1(\zeta)$ is the standard exponential integral function,

$$E_1(\zeta) \equiv \int_\zeta^\infty \frac{e^{-u}}{u} du . \tag{65}$$

Analogous integral expressions for the moment $\Sigma_d^1(\rho)$ and $\Sigma_d^4(\rho)$ are presented in Appendix D. Once again, asymptotic evaluation of these integrals recovers the same $\sim 1/\rho$ and $\sim \log(1/\rho)$ scaling observed in our earlier heuristic

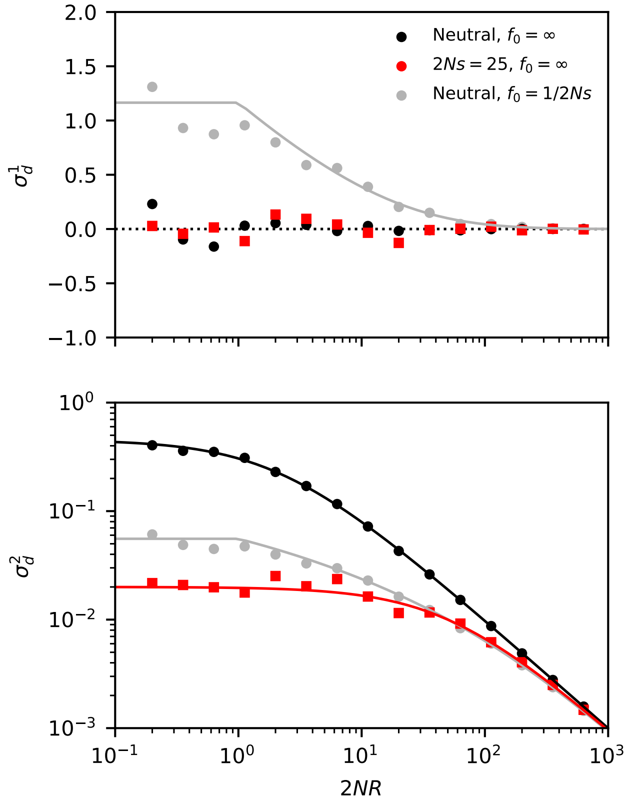

FIG. 5 **LD contains residual signatures of purifying selection after controlling for mutation frequencies.** The top and bottom panels compare the first (top) and second (bottom) LD moments for a pair of neutral (black) and strongly deleterious mutations (red) across a range of recombination rates. Symbols denote the results of forward-time simulations, and the solid lines denote the theoretical predictions from Eq. (2) (black) and Eq. (66) (red). The grey symbols show frequency-weighted neutral mutations with the same present-day frequency spectrum as the deleterious mutations. For small recombination rates, the neutral control group displays an excess of coupling linkage ($\sigma_d^1 > 0$) driven by ancient nested mutations, which are suppressed in the deleterious case.

analysis for large and small ρ, respectively, while the numerical solutions accurately capture the quantitative behavior across the full range of ρ (Fig. 4). This full solution is often quite useful in practice, since the convergence to the asymptotic limits can be rather slow. In particular, we see that the corrections to the small ρ limit scale as $\sim 1/\log(1/\rho)$, which implies that extremely small values of ρ (as low as $\sim 10^{-5}$) are required to achieve good numerical accuracy. This leaves a broad intermediate regime ($10^{-5} \lesssim \rho \lesssim 10$) in which Eq. (63) is critical for enabling quantitative comparisons with data.

Strong selection or recombination. The final limit we will consider is one in which at least one of scaled selection coefficients (γ_i) or the scaled recombination rate (ρ) is large compared to one. Physically, this approximation means that genetic drift is weak compared to the forces of selection and/or recombination. In this limit, it is possible to solve for the characteristic curve in Eq. (56b) using a separation of timescales approach, treating the ψ^2 term as a perturbative correction. This perturbation expansion is outlined in Appendix E. We find that the first two moments are given by

$$F_1 \approx 0 \, , \quad F_2 \approx \frac{2 + \gamma_A + \gamma_B + \rho}{(\gamma_A + \gamma_B + \gamma_\epsilon + \rho)^2} \, , \tag{66}$$

which matches the asymptotic behavior in the non-recombining ($\rho = 0$) and neutral ($\gamma_i = 0$) limits above. By comparing this result with the neutral version in Eq. (63), we observe a quantitative confirmation of our heuristic prediction that LD is lower among deleterious mutations than among neutral mutations with identical present day frequencies ($f_{0,\text{eff}} = 1/2Ns$). The most pronounced difference occurs for the first moment σ_d^1, where frequency-matched neutral mutations display an excess of coupling linkage ($\sigma_d^1 > 0$) compared to the deleterious case ($\sigma_d^1 \approx 0$), which can be observed for recombination rates as large as $\rho \approx 50$ (Fig. 5).

Transition to Quasi-Linkage Equilibrium (QLE)

The perturbation expansion in Eq. (56) is valid to lowest order in $f_0 \ll 1$, which means that it cannot capture the transition to the quasi-linkage equilibrium (QLE) regime that occurs when $\rho f_0 {\sim} 1$. Nevertheless, we can obtain an analogous set of predictions for this regime by returning to the underlying stochastic differential equations in Eq. (8), and focusing on cases where $\rho \gg 1, \gamma_i$ and $f_0 \ll 1$, but where ρf_0 is not necessarily small compared to one. In this limit, Eq. (8) can be solved using a separation of timescales approach, in which f_{AB} evolves on a fast timescale,

$$\frac{\partial f_{AB}}{\partial t} = R f_{Ab} f_{aB} - R f_{AB} + \sqrt{\frac{f_{AB}}{N}} \eta(t) \, , \tag{67}$$

while f_{Ab} and f_{aB} are effectively fixed. These single-mutant frequencies then evolve on longer timescales according to the single-locus dynamics in Eq. (54). In this approximation, the fast dynamics in Eq. (67) approach an instantaneous equilibrium,

$$p(f_{AB} | f_{Ab}, f_{aB}) \propto f_{AB}^{2NR f_{Ab} f_{aB} - 1} e^{-2NR f_{AB}} \, , \tag{68}$$

on a timecale of order $\sim 1/R$, which is much shorter than the fluctuation timescales of f_{Ab} and f_{aB} in the limit that $\rho \gg 1$. This allows us to easily calculate various LD statistics from the conditional moments of Eq. (68), by averaging over the single-locus distributions of f_{Ab} and

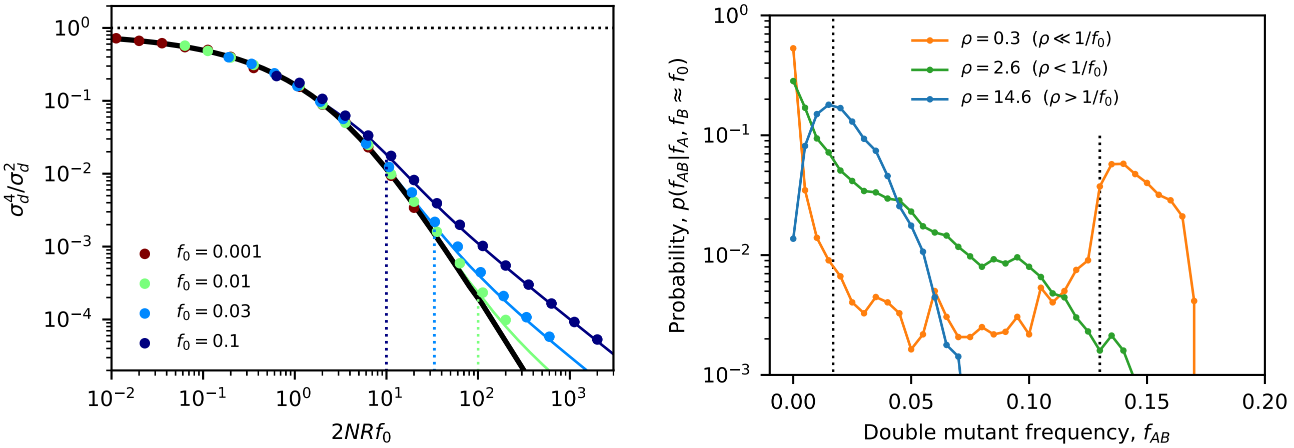

FIG. 6 **Higher-order fluctuations reveal the transition to quasi-linkage equilibrium.** Left panel: an analogous version of the neutral collapse plot in Fig. 4 for the higher-order LD moment $\sigma_d^4(f_0)$. Symbols denote the results of forward time simulations for a range of recombination rates, which are colored by the corresponding value of f_0. The solid black line shows the prediction from the perturbation expansion in Eqs. (D24) and (D25), and the dashed lines indicate the position, $NRf_0 \sim 1/f_0$, where the perturbation expansion is predicted to break down. The solid colored lines show the asymptotically matched predictions from Eq. (F6), which capture the transition to the quasi-linkage equilibrium regime. Right panel: the conditional distribution of the double mutant frequency for fixed values of the marginal mutation frequencies, $f_A \approx f_B \approx f_0$. Colored lines show forward-time simulations for pairs of neutral mutations, in which the marginal frequencies of both mutations were observed in the range $0.13 \leq f_A, f_B \leq 0.17$; the double mutant frequency was further downsampled to $n = 200$ individuals to enhance visualization. The dashed lines indicate the approximate positions of linkage equilibrium ($f_{AB} \approx f_0^2$, left) and perfect linkage ($f_{AB} \approx f_0$; right). Conditional distributions are shown for three different recombination rates, whose characteristic shapes illustrate the transition between the mutation-dominated ($NRf_0 \ll 1$; orange), clonal recombinant ($1 \ll NRf_0 \ll 1/f_0$; green) and quasi-linkage equilibrium ($NRf_0 \gg 1/f_0$; blue) regimes.

f_{aB} in Eq. (55). For the first few moments of $\sigma_d^k(f_0)$, we find that

$$\sigma_d^1(f_0) = 0\,, \quad \sigma_d^2(f_0) = \frac{1}{2NR}\,, \\ \sigma_d^4(f_0) = \frac{6 + 6NRf_0^2}{(2NR)^3}\,. \tag{69}$$

This confirms our heuristic result that the first two moments of σ_d^k are identical in the clonal recombinant and quasi-linkage equilibrium regimes, while the higher moments start to diverge due to the differences in the statistical fluctuations of f_{AB}. These differences can be observed by examining the ratio

$$\eta(f_0) = \frac{\sigma_d^4(f_0)}{3\sigma_d^2(f_0)}\,, \tag{70}$$

which shifts from a rapid $\sim 1/(NR)^2$ decay in the clonal recombinant regime to a shallower $\sim f_0^2/NR$ decay in the QLE regime, while saturating to a constant value when $NRf_0 \ll 1$ (Fig. 6, left).

The differences between these regimes are even easier to observe by examining the full conditional distribution of the double mutant frequency, $p(f_{AB}|f_A, f_B{\approx}f_0)$, at a fixed value of the marginal mutation frequencies, $f_A \approx f_B \approx f_0$ (Fig. 6, right). In the QLE regime, Eq. (68) shows that the conditional distribution develops a peak around the linkage equilibrium value $f_{AB} \approx f_A f_B \approx f_0^2$, while the clonal recombinant regime has a much broader distribution with a mode at $f_{AB} = 0$ and an exponential cutoff at $f_{AB}{\sim}1/NR$. These characteristic shapes are qualitatively distinct from the conditional distributions that are observed in the mutation-dominated regime ($NRf_0 \ll 1$), which have a bimodal shape with one peak at $f_{AB} = 0$ and a smaller peak at $f_{AB} \approx f_A \approx f_B \approx f_0$. This suggests that the shape of the conditional distribution $p(f_{AB}|f_A, f_B \approx f_0)$ might provide a particularly robust test for distinguishing between different rates of recombination. We will return to this topic in the Discussion when we discuss potential applications of our results to genomic data from bacteria.

Estimating frequency-resolved LD in finite samples

So far, our formal analysis has focused on predicting ensemble averages of various LD statistics at a single pair of genetic loci. To connect these results with empirical data, we will often want to estimate these ensemble averages in a slightly different way, by summing over many functionally similar pairs of genetic loci observed in a finite sample of n genomes. In these cases, we will not be able to observe the haplotype frequencies that enter into $\sigma_d^k(f_0)$ directly, but must instead infer them from the discrete counts, n_{AB}, n_{AB}, n_{aB}, and n_{ab} that are observed in our finite sample. We will assume that these haplotype counts are randomly sampled from the underlying population, so that they are multinomially distributed around the current haplotype frequencies:

$$\Pr[\vec{n}|\vec{f}] = \frac{n! \cdot f_{Ab}^{n_{Ab}} \cdot f_{aB}^{n_{aB}} \cdot f_{AB}^{n_{AB}} \cdot f_{ab}^{n_{ab}}}{n_{Ab}! \cdot n_{aB}! \cdot n_{AB}! \cdot n_{ab}!} \,. \tag{71}$$

In sufficiently large samples ($n \to \infty$), the haplotype counts will remain close to their expected values $n_i \approx nf_i$, and $\sigma_d^k(f_0)$ can be well-approximated by setting $f_i = n_i/n$ in Eq. (7). However, for sufficiently low frequencies ($nf_i{\sim}10$), the additional uncertainty in f_i will cause the naive estimator to be biased away from the true value of $\sigma_d^k(f_0)$. Our heuristic results show that these low frequencies will generically dominate LD estimates — even in large samples — for sufficiently large values of NR or Ns_i, or for sufficiently low choices of f_0. Unbiased estimators of $\sigma_d^k(f_0)$ are therefore essential for extrapolating across the full range of frequency scales.

In this section, we will develop one particular class of estimators by generalizing an approach that we and others have previously used to estimate the unweighted version of σ_d^2 in Eq. (4) (Garud et al., 2019; Ragsdale and Gravel, 2020). To extend this result to the frequency-resolved case, we will first take advantage of the fact that the multinomial distribution in Eq. (71) reduces to a product of independent Poisson distributions,

$$\begin{aligned}\Pr[\vec{n}|\vec{f}] \approx \frac{(nf_{Ab})^{n_{Ab}}}{n_{Ab}!} e^{-nf_{Ab}} \cdot \frac{(nf_{aB})^{n_{aB}}}{n_{aB}!} e^{-nf_{aB}} \\ \times \frac{(nf_{AB})^{n_{AB}}}{n_{AB}!} e^{-nf_{AB}} \,,\end{aligned} \tag{72}$$

in the limit that mutations are rare ($f_A, f_B \ll 1$). This joint distribution admits a general moment formula,

$$\begin{aligned}\left\langle \frac{n_{Ab}! n_{aB}! n_{AB}! e^{-x(n_{Ab}-i)-y(n_{aB}-j)-z(n_{AB}-k)}}{n^{i+j+k}(n_{Ab}-i)!(n_{aB}-j)!(n_{AB}-k)!} \middle| \, \vec{f} \right\rangle = \\ f_{Ab}^i f_{aB}^j f_{AB}^k e^{-nf_{aB}(1-e^{-x})-nf_{aB}(1-e^{-y})-nf_{AB}(1-e^{-z})} \,,\end{aligned} \tag{73}$$

for arbitrary integers i,j, and k, and arbitrary real numbers x, y, and z. Thus, for the special choice

$$x^* \equiv y^* \equiv \log\left(\frac{nf_0}{nf_0 - 1}\right), \quad z^* \equiv \log\left(\frac{nf_0}{nf_0 - 2}\right), \tag{74}$$

the conditional expectation reduces to

$$\begin{aligned}\left\langle \frac{n_{Ab}! n_{aB}! n_{AB}! e^{-x^*(n_{Ab}-i)-y^*(n_{aB}-j)-z^*(n_{AB}-k)}}{n^{i+j+k}(n_{Ab}-i-1)!(n_{aB}-j-1)!(n_{AB}-k-1)!} \middle| \, \vec{f} \right\rangle \\ = f_{Ab}^i f_{aB}^j f_{AB}^k e^{-f_A/f_0 - f_B/f_0} \,.\end{aligned} \tag{75}$$

This motivates us to define the function,

$$\begin{aligned}M_{i,j,k}(\vec{n}) = \Bigg[\frac{n_{Ab}! \left(1 - \frac{1}{nf_0}\right)^{n_{Ab}-i}}{n^i (n_{Ab}-i)!} \\ \times \frac{n_{aB}! \left(1 - \frac{1}{nf_0}\right)^{n_{aB}-i}}{n^j (n_{aB}-j)!} \cdot \frac{n_{AB}! \left(1 - \frac{2}{nf_0}\right)^{n_{AB}-k}}{n^k (n_{AB}-k)!} \Bigg] \,,\end{aligned} \tag{76}$$

whose total expectation — which now averages over sampling noise in addition to the underlying evolutionary stochasticity — satisfies the identity

$$\langle M_{i,j,k}(\vec{n}) \rangle = \left\langle f_{Ab}^i f_{aB}^j f_{AB}^k e^{-f_A/f_0 - f_B/f_0} \right\rangle \,. \tag{77}$$

Thus, we see that for this special choice of exponential weighting function, there is a simple relationship between the ensemble averages of haplotype frequencies and genome-wide averages over haplotype counts. Using this formula, it is a straightforward (though tedious) task to derive a corresponding set of estimators for $\sigma_d^k(f_0)$, by expanding the f_A and f_B terms in Eq. (7) and iteratively applying Eq. (77). Expressions for the first few moments of $\sigma_d^k(f_0)$ are listed in Appendix G.

Applications to synonymous and nonsynonymous LD

LD curves are frequently calculated for pairs of synonymous or nonsynonymous mutations separated by similar coordinate distances ℓ (or ideally, by similar map lengths R). In these cases, the empirical estimators in the previous section converge to a weighted average,

$$\sigma_{d,i}^k(R, f_0) \approx \frac{\iiint \sigma_d^k(s_A, s_B, \epsilon, R, f_0) \left(\frac{1}{1+2Ns_Af_0}\right) \left(\frac{1}{1+2Ns_Bf_0}\right) \rho_i(s_A)\rho_i(s_B)\rho_i^\epsilon(\epsilon|s_A, s_B)\, ds_A\, ds_B\, d\epsilon}{\iiint \left(\frac{1}{1+2Ns_Af_0}\right) \left(\frac{1}{1+2Ns_Bf_0}\right) \rho_i(s_A)\rho_i(s_B)\, ds_A\, ds_B} \,, \tag{78}$$

where $\rho_i(s)$ denotes the distribution of fitness costs of synonymous ($i = S$) or nonsynonymous ($i = N$) mutations, and $\rho_i^\epsilon(\epsilon|s_A, s_B)$ denotes the corresponding distribution of epistatic interactions. In this way, any differences between the observed $\sigma_{d,N}^k$ and $\sigma_{d,S}^k$ curves can provide additional information about the differences in their underlying fitness costs.

To understand the consequences of the average in Eq. (78), recall that our earlier analytical expressions showed that deleterious fitness costs generally lead to lower values of $\sigma_d^k(s_A, s_B, \epsilon, R, f_0)$, where the magnitude of this effect depends on the relative values of R and f_0. The additional factors that appear in the average in Eq. (78) will further downweight the contributions of mutations with costs larger than $\sim 1/Nf_0$. This shows that strongly deleterious mutations ($s \gg 1/Nf_0$) will have a negligible impact in the numerator of Eq. (78), as long as there is an appreciable fraction of mutations with smaller fitness costs. However, for the same reasons, these strongly deleterious mutations will also have a negligible impact on the denominator of Eq. (78), which implies that they will have a negligible overall contribution to the site-averaged LD statistics $\sigma_{d,N}^k(R, f_0)$ and $\sigma_{d,S}^k(R, f_0)$.

At the same time, we have seen that very weakly deleterious mutations ($s_{AB} \ll 1/Nf_0$) produce σ_d^k values that are nearly indistinguishable from neutral mutations, differing only by a slowly varying $\sim \log(1/Ns_{AB}f_0)$ factor. These mutations contribute to the averages in $\sigma_{d,N}^k(R, f_0)$ and $\sigma_{d,S}^k(R, f_0)$, but they cannot contribute to differences between the two quantities. Thus, in the absence of epistasis, we expect that the differences between synonymous and nonsynonymous LD will be driven by a narrow range of mutations with fitness costs $\mathcal{O}(1/Nf_0)$, and will mainly be visible when $NRf_0 \ll 1$.

On one hand, this sensitivity suggests that it might be possible to infer detailed information about $\rho_N(s)$ by comparing $\sigma_N^k(f_0)$ and $\sigma_S^k(f_0)$ values across a range of frequency scales. On the other hand, our analytical expressions show that these marginal fitness costs will only lead to $\mathcal{O}(1)$ differences in $\sigma_N^k(f_0)$, which will sensitively depend on the precise value of the integral in Eq. (78). We leave a more detailed exploration of this dependence for future work. We also note that this limited resolution is no longer an issue in the presence of epistasis: strong synergistic epistasis between weakly selected mutations can produce large changes in $\sigma_N^k(f_0)$ if they are sufficiently common.

DISCUSSION

Contemporary patterns of linkage disequilibrium contain important information about the evolutionary forces at work within a population, which shape genetic variation over a vast range of length and time scales. Here, we have introduced a forward-time framework for predicting linkage disequilibrium between pairs of neutral or deleterious mutations as a function of their present-day frequency scale f_0. This additional dependence turned out to be more than a statistical curiosity, but instead enabled new insights into the dynamics of linkage disequilibrium that had been difficult to obtain from existing methods (McVean, 2002; Ohta and Kimura, 1971; Song and Song, 2007).

Our frequency-resolved approach shares some common features with existing moment-based approaches (Good and Desai, 2013; Ragsdale and Gravel, 2019; Song and Song, 2007; ?; ?), which have developed recursion relations to calculate arbitrary higher-order moments of D, f_A, and f_B. These different moments also emphasize mutations with different frequencies, and can in principle be combined to single out particular frequency ranges as we have done above. For example, by Taylor expanding the exponential in Eq. (7), our frequency-weighted statistic $\sigma_d^k(f_0)$ can be expressed as an infinite sum over the higher-order moments $\langle D^k f_A^p f_B^q \rangle$. By restricting our attention to rare mutations ($f_0 \ll 1$), our present approach is able to sum up these infinite contributions analytically, without requiring the truncation schemes or moment closure approximations employed by previous methods. These benefits are particularly useful at the lowest frequency scales ($f_0 \ll 1$), which are dominated by increasingly higher-order terms in the formal series expansion.

Our focus on rare mutations also allowed us to obtain a simple heuristic picture of linkage disequilibrium that emphasizes the underlying dynamics of the lineages involved (Fig. 2). We saw that the frequency scale f_0 can dramatically influence these dynamics, in a way that primarily depends on frequency-rescaled quantities like NRf_0 and Nsf_0. Our lineage-based picture highlighted the crucial importance of ancient nested mutations (Fig. 2C), which are substantially older than typical segregating variants, but which provide an increasingly large contribution to LD among tightly linked loci ($NRf_0 \lesssim 1$) with neutral or weakly deleterious fitness costs ($Nsf_0 \lesssim 1$). In these cases, we saw that ancient nested mutations will create an excess of coupling linkage ($f_{AB} > f_A f_B$) that qualitatively resembles the effects of antagonistic epistasis. This excess coupling linkage has previously been observed in computer simulations and in genomic data from diverse organisms (Garcia and Lohmueller, 2020; Sandler et al., 2020; Sohail et al., 2017), where it has fueled an ongoing debate about the inference of epistasis from patterns of nonsynonymous and synonymous LD in a variety of species. Our analytical calculations suggest a potential mechanism for this counterintuitive behavior, and they demonstrate that this effect will generically arise even in the absence of admixture or other complex demographic scenarios.

Our results also allow us to answer a question we posed

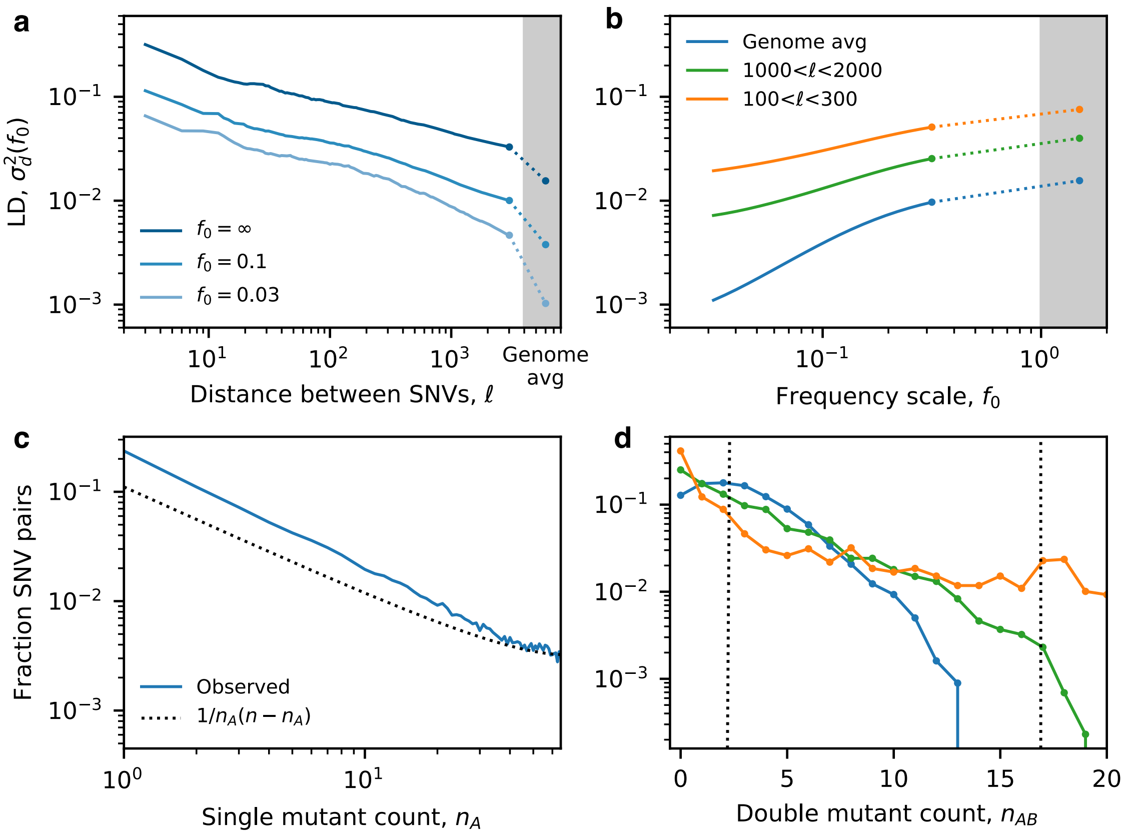

FIG. 7 **Frequency-resolved LD in the commensal human gut bacterium *Eubacterium rectale*.** Single nucleotide variant (SNVs) were obtained for a sample of $n = 109$ unrelated strains reconstructed from different human hosts (Garud et al., 2019) (Appendix H). (a) Frequency-weighted LD ($\sigma_d^2(f_0)$) as a function of coordinate distance (ℓ) between 4-fold degenerate synonymous SNVs in core genes. Solid lines were obtained by applying the unbiased estimator in Appendix G to all pairs of SNVs within 0.2 log units of ℓ, while the points depict genome-wide averages calculated from randomly sampled pairs of SNVs from widely separated genes. The two estimates are connected by a dashed line for visualization. (b) Analogous $\sigma_d^2(f_0)$ curves as a function of the frequency scale f_0. (c) The single site frequency spectrum, estimated from the fraction of SNV pairs in which the first mutation is observed with a given minor allele count, n_A. (d) The conditional distribution of the double mutant frequency for fixed values of the marginal mutation frequencies, $f_A \approx f_B \approx f_0$. Colored lines show the observed distributions for pairs of SNVs with marginal mutation frequencies in the range $0.13 \leq f_A, f_B \leq 0.17$; dashed lines indicate approximate positions of linkage equilibrium ($f_{AB} \approx f_0^2$, left) and perfect linkage ($f_{AB} \approx f_0$; right). The shapes of the three distributions are qualitatively similar to the the mutation-dominated ($NRf_0 \ll 1$; orange), clonal recombinant ($1 \ll NRf_0 \ll 1/f_0$; green) and quasi-linkage equilibrium ($NRf_0 \gg 1/f_0$; blue) regimes predicted in Fig. 6.

at the beginning of this work: do differences between synonymous and nonsynonymous LD curves primarily arise from differences in their underlying mutation frequencies? Our analytical results demonstrate that this is not the case: the key difference is that strongly deleterious mutations ($Nsf_0 \gg 1$) can no longer sustain the ancient nested mutations in Fig. 2C, leading to lower levels of LD compared to neutral mutations with similar present-day frequencies ($f_0 \sim 1/Ns$; Fig. 5). This shows that ordinary negative selection can generate differences between synonymous and nonsynonymous LD that qualitatively resemble the effects of negative epistasis. This could be an important potential confounder for efforts to infer negative epistasis by comparing levels of nonsynonymous and synonymous LD (Sandler et al., 2020; Sohail et al., 2017). However, we also saw that these strongly deleterious mutations can have a negligible impact on certain genome-wide averages like σ_d^2, due to their lower marginal frequencies. Thus, the quantitative magnitude of this effect can strongly depend on the underlying distribution of fitness effects as well as the averaging scheme employed. Moreover, while additive fitness costs can lead to lower *relative* values of LD, we also saw that they cannot produce negative values of $\sigma_d^1(f_0)$ on their own. This suggests that negative genome-wide values of signed LD may constitute a more robust indicator of negative epistasis than relative reductions in LD over synonymous sites.

More generally, our results provide a framework for leveraging the increasingly large sample sizes of modern genomic datasets to quantify the scaling behavior of linkage disequilibrium across a range of underlying frequency scales. These scaling behaviors have a long history of application in other areas of statistical physics (Meshulam et al., 2019; Stanley, 1999), and are commonly used in population genetics to infer evolutionary parameters from the shape of the single-site frequency spectrum (Lawrie and Petrov, 2014; Ragsdale et al., 2018). Our results provide a framework for extending this approach to multi-site statistics like linkage disequilibrium, potentially creating new opportunities to probe the underlying recombination process across a wide range of genomic length and time scales.

An example of this approach is illustrated in Fig. 7, which calculates frequency-resolved LD curves for 109 worldwide strains of the commensal human gut bacterium *Eubacterium rectale* (Appendix H). In a previous study, my collaborators and I used this dataset to infer the presence of widespread homologous recombination in the global population of *E. rectale*, by examining how the unweighted version of σ_d^2 decays as a function of the coordinate distance ℓ (Garud et al., 2019). Our new frequency-resolved estimators now provide an analogous manifold of LD curves, $\sigma_d^2(\ell, f_0)$, which allow us to examine the dynamics of LD across nearly two decades of frequency space (Fig. 7A,B). At a qualitative level, these empirical curves are similar to their theoretical counterparts above (Figs 4 and 5), with smaller mutation frequencies and/or longer coordinate distances leading to lower values of LD. However, the quantitative dependence on ℓ and f_0 indicates dramatic departures from the simplest neutral null models analyzed in this work. In particular, the *E. rectale* data suggest that larger coordinate distances are more sensitive to reductions in f_0 (Fig. 7A,B), while our theoretical models predict the opposite trend (Fig. 5). Moreover, this unusual frequency dependence is observed even at the largest coordinate distances ($\ell{\sim}10^6$bp), where the overall reduction in σ_d^2 might normally suggest convergence to the recombination-dominated regime ($NR \gg 1$). This example shows how analytical predictions of frequency-resolved LD statistics can help identify surprising features of the data that warrant future study, yet would be difficult to identify from intuition alone.

In this case, the divergence between theory and data might have been anticipated, given that the marginal mutation frequencies in *E. rectale* already deviate from the $\sim 1/f$ dependence predicted under the simplest neutral null model (Fig. 7C). It is possible that the modest enrichment of rare mutations observed in the data could bias the relevant averages below the nominal value of f_0, leading to somewhat stronger realized levels of linkage than would be expected under the simplest versions of our model. To reduce these potential uncertainties induced by the average in $\sigma_d^2(f_0)$, it is also useful to quantify the same LD patterns in a different way, by examining the conditional distribution of the double mutant frequency for a *specific* value of the marginal mutation frequencies, $f_A \approx f_B \approx f_0$ (Fig. 7D). When $f_0{\approx}10\%$, the data display a clear transition between the three characteristic regimes identified in Fig. 6, with quasi-linkage equilibrium emerging at the largest coordinate distances ($\ell{\sim}10^6$) and mutation-dominated behavior at shorter genetic distances ($100 < \ell < 300$). On intermediate length scales ($\ell{\sim}1000$), the LD distribution transitions to an exponential shape expected in the clonal recombinant regime ($f_0^{-1} \ll NR(\ell) \ll f_0^{-2}$), which provides a direct empirical demonstration of this qualitatively new behavior. The boundaries between these regimes provide an independent set of bounds on the corresponding recombination rates,

$$10 \lesssim NR(\ell{\approx}10^3) \lesssim 100 \lesssim NR(\ell{\approx}10^6)\,, \qquad (79)$$

which no longer require extrapolation over multiple distance scales, or any associated assumptions about the shape of the $R(\ell)$ curve. This example shows how the sampling distributions of different LD statistics can provide new insights into the dynamics of the underlying recombination process.

Of course, these empirical comparisons should be treated with a degree of caution, since our theoretical analysis focused on an extremely simple null model that lacks many of the complexities associated with real microbial populations. Our results suggest that it would be interesting to extend these approaches to account for other factors that might be relevant at short time scales, including time-varying population sizes, linked selection, and certain forms of spatial structure. We believe that our lineage-based framework will provide a useful starting point for predicting the dynamics of linkage disequilibrium across these diverse evolutionary scenarios, which would allow us to better exploit the unique features of modern genomic datasets.

ACKNOWLEDGMENTS

I thank Nandita Garud and Daniel Fisher for useful discussions, and Anastasia Lyulina and Zhiru Liu for valuable feedback on the manuscript. This work was supported in part by the Alfred P. Sloan Foundation (FG-2021-15708), a Terman Fellowship from Stanford University, and the Miller Institute for Basic Research in Science at the University of California, Berkeley.

REFERENCES

Allix-Béguec, C., Arandjelovic, I., Bi, L., Clifton, D., Crook, D., Fowler, P., Gibertoni Cruz, A., Golubchik, Hoosdally, S., Hunt, M., Iqbal, Z., Lipworth, S., Peto, T., Thwaites, G., Walker, A., Walker, T., Wilson, D., Wyllie, D., and Yang, Y. 2018. Prediction of susceptibility to first-line tuberculosis drugs by dna sequencing. *New England Journal of Medicine* 379:1403–1415.

Ansari, M. A. and Didelot, X. 2014. Inference of the properties of the recombination process from whole bacterial genomes. *Genetics* 196:253–265.

Arnold, B., Sohail, M., Wadsworth, C., Corander, J., Hanage, W. P., Sunyaev, S., and Grad, Y. H. 2020. Fine-scale haplotype structure reveals strong signatures of positive selection in a recombining bacterial pathogen. *Molecular Biology and Evolution* 37:417–428.

Chakravarti, A., Buetow, K. H., Antonarakis, S., Waber, P., Boehm, C., and Kazazian, H. 1984. Nonuniform recombination within the human beta-globin gene cluster. *American journal of human genetics* 36:1239.

Coop, G. and Ralph, P. 2012. Patterns of neutral diversity under general models of selective sweeps. *Genetics* 192:205–224.

Cvijović, I., Good, B. H., and Desai, M. M. 2018. The effect of strong purifying selection on genetic diversity. *Genetics* 209:1235–1278.

Desai, M. M. and Fisher, D. S. 2007. Beneficial mutation selection balance and the effect of genetic linkage on positive selection. *Genetics* 176:1759–1798.

Eberle, M. A., Rieder, M. J., Kruglyak, L., and Nickerson, D. A. 2006. Allele frequency matching between snps reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLoS genetics* 2.

Ewens, W. J. 2004. Mathematical Population Genetics. Springer-Verlag, New York, second edition.

Fisher, D. S. 2007. Evolutionary dynamics, pp. 395–446. *In* M. M. Jean-Philippe Bouchaud and J. Dalibard (eds.), Complex Systems, volume 85 of *Les Houches*. Elsevier.

Garcia, J. A. and Lohmueller, K. E. 2020. Negative linkage disequilibrium between amino acid changing variants reveals interference among deleterious mutations in the human genome. *bioRxiv* .

Gardiner, C. 1985. Handbook of Stochastic Methods. Springer, New York.

Garud, N. R., Good, B. H., Hallatschek, O., and Pollard, K. S. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS biology* 17:e3000102.

Garud, N. R., Messer, P. W., Buzbas, E. O., and Petrov, D. A. 2015. Recent selective sweeps in north american drosophila melanogaster show signatures of soft sweeps. *PLoS genetics* 11.

Good, B. H. 2016. Molecular Evolution in Rapidly Evolving Populations. PhD thesis, Harvard University, Cambridge MA.

Good, B. H. and Desai, M. M. 2013. Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theor Pop Biol* 85:86–102.

Harris, K. and Nielsen, R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet* 9:e1003521.

Hedrick, P. W. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331–341.

Hill, W. and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theoretical and applied genetics* 38:226–231.

Kamm, J. A., Terhorst, J., and Song, Y. S. 2017. Efficient computation of the joint sample frequency spectra for multiple populations. *Journal of Computational and Graphical Statistics* 26:182–194.

Kang, J. T. and Rosenberg, N. A. 2019. Mathematical properties of linkage disequilibrium statistics defined by normalization of the coefficient d= pab–papb. *Human Heredity* 84:127–143.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443.

Kim, Y. and Nielsen, R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167:1513–1524.

Kimura, M. 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1:177–232.

Kimura, M. and Ohta, T. 1973. The age of a neutral mutant persisting in a finite population. *Genetics* 75:199–212.

Lawrie, D. S. and Petrov, D. A. 2014. Comparative population genomics: power and principles for the inference of functionality. *Trends in Genetics* 30:133 – 139.

Lewontin, R. 1964. The interaction of selection and linkage. i. general considerations; heterotic models. *Genetics* 49:49.

Lewontin, R. 1988. On measures of gametic disequilibrium. *Genetics* 120:849–852.

Li, H. and Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.

Lynch, M., Xu, S., Maruki, T., Jiang, X., Pfaffelhuber, P., and Haubold, B. 2014. Genome-wide linkage-disequilibrium profiles from single individuals. *Genetics* 198:269–281.

McVean, G. 2007. The structure of linkage disequilibrium around a selective sweep. *Genetics* 175:1395–1406.

McVean, G. A. 2002. A genealogical interpretation of linkage disequilibrium. *Genetics* 162:987–991.

McVean, G. A., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* 304:581–584.

Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W., and Bialek, W. 2019. Coarse graining, fixed points, and scaling in a large population of neurons. *Phys. Rev. Lett.* 123:178103.

Neher, R. A. and Hallatschek, O. 2013. Genealogies in rapidly adapting populations. *Proc Nat Acad Sci* 110:437–442.

Ohta, T. and Kimura, M. 1971. Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* 68:571.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 176:649–662.

Petit III, R. A. and Read, T. D. 2018. Staphylococcus aureus viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261.

Pfaffelhuber, P., Lehnert, A., and Stephan, W. 2008. Linkage disequilibrium under genetic hitchhiking in finite populations. *Genetics* 179:527–537.

Pokalyuk, C. 2012. The effect of recurrent mutation on the linkage disequilibrium under a selective sweep. *Journal of mathematical biology* 64:291–317.

Polanski, A. and Kimmel, M. 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics* 165:427–436.

Ragsdale, A. P. and Gravel, S. 2019. Models of archaic admixture and recent history from two-locus statistics. *PLoS genetics* 15:e1008204.

Ragsdale, A. P. and Gravel, S. 2020. Unbiased estimation of linkage disequilibrium from unphased data. *Molecular Biology and Evolution* 37:923–932.

Ragsdale, A. P., Moreau, C., and Gravel, S. 2018. Genomic inference using diffusion models and the allele frequency spectrum. *Current Opinion in Genetics & Development* 53:140 – 147. Genetics of Human Origins.

Rogers, A. R. 2014. How population growth affects linkage disequilibrium. *Genetics* 197:1329–1341.

Rosen, M. J., Davison, M., Bhaya, D., and Fisher, D. S. 2015. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science* 348:1019–1023.

Rosen, M. J., Davison, M., Fisher, D. S., and Bhaya, D. 2018. Probing the ecological and evolutionary history of a thermophilic cyanobacterial population via statistical properties of its microdiversity. *PloS one* 13.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.

Sandler, G., Wright, S. I., and Agrawal, A. F. 2020. Using patterns of signed linkage disequilibria to test for epistasis in flies and plants. *bioRxiv* .

Sawyer, S. A. and Hartl, D. L. 1992. Population genetics of polymorphism and divergence. *Genetics* 132:1161–1176.

Shu, Y. and McCauley, J. 2017. Gisaid: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22:30494.

Slatkin, M. 2008. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* 9:477–485.

Sohail, M., Vakhrusheva, O. A., Sul, J. H., Pulit, S. L., Francioli, L. C., van den Berg, L. H., Veldink, J. H., de Bakker, P. I., Bazykin, G. A., Kondrashov, A. S., et al. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. *Science* 356:539–542.

Song, Y. S. and Song, J. S. 2007. Analytic computation of the expectation of the linkage disequilibrium coefficient r2. *Theoretical population biology* 71:49–60.

Stanley, H. E. 1999. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.* 71:S358–S366.

Stephan, W., Song, Y. S., and Langley, C. H. 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* 172:2647–2663.

VanLiere, J. M. and Rosenberg, N. A. 2008. Mathematical properties of the r2 measure of linkage disequilibrium. *Theoretical population biology* 74:130–137.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. 2017. 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics* 101:5–22.

Weissman, D. B., Feldman, M. W., and Fisher, D. S. 2010. The rate of fitness-valley crossing in sexual populations. *Genetics* 186:1389–1410.

Weissman, D. W., Desai, M. M., Fisher, D. S., and Feldman, M. W. 2009. The rate at which asexual populations cross fitness valleys. *Theoretical Population Biology* 75:286–300.

Appendix A: Forward-time simulations

Forward-time simulations were used to compare our analytical predictions to the full two-locus model in Eq. (3). These simulations used a discrete generation, Wright-Fisher sampling scheme, in which the number of individuals with each haplotype at generation $t+1$ was drawn from a Poisson distribution with mean equal to the expected number of individuals predicted by Eq. (3), given the haplotype frequencies at time t. To enhance computational efficiency for calculating LD statistics, our program only simulated timepoints in which both mutations were segregating in the population at the same time. This scheme was implemented by first drawing an initial (single-mutant) haplotype from the single-locus site frequency spectrum (Sawyer and Hartl, 1992),

$$p(f|s) \propto \frac{e^{2Ns(1-f)}-1}{x(1-x)}\,, \tag{A1}$$

and then introducing a second mutation in a single individual from either the mutant or wildtype background with probability f or $1-f$, respectively. The resulting population was then evolved until one of the mutations went extinct, and the process was restarted with a new pair of mutations. The frequencies of the four haplotypes were recorded every Δt generations, and were used to generate the figures in the main text. The simulations in this work were performed with a population size of $N = 10^5$ and a sampling interval of $\Delta t = 100$.

Appendix B: Perturbative solution of the generating function for small f_0

To obtain the perturbative solution of Eq. (50) listed in Eq. (59), it is be helpful to define a zeroth order differential operator,

$$\mathcal{L} = \frac{\partial}{\partial \tau} + \left(\gamma_A x + x^2\right)\frac{\partial}{\partial x} + \left(\gamma_B y + y^2\right)\frac{\partial}{\partial y} + \left(\gamma_{AB} z + z^2 - \rho(x+y)\right)\frac{\partial}{\partial z} \tag{B1}$$

so that Eq. (50) can be written as

$$\mathcal{L}H = -\theta(x+y)H + \theta f_0 z\left(\frac{\partial H}{\partial x} + \frac{\partial H}{\partial y}\right) - \rho f_0 (z{-}x{-}y)\frac{\partial^2 H}{\partial x \partial y} \tag{B2}$$

with all of the θ and f_0 dependence is confined to the right-hand side. Substituting the series expansion in Eq. (52) into this equation and grouping like terms, we find that the zeroth order solution is $H \approx 1$, as expected. The first order correction, which enters at $\mathcal{O}(\theta)$, satisfies the equation

$$\mathcal{L}H_{1,0} = -(x+y)\,, \tag{B3}$$

which can be solved using the method of characteristics. To see this, we define a function

$$\Phi_{1,0}(\tau_R) = H_{1,0}(x(\tau_R), y(\tau_R), \tau - \tau_R)\,, \tag{B4}$$

where the characteristic curves $x(\tau_R)$ and $y(\tau_R)$ are given by

$$\begin{aligned}
\partial_{\tau_R} x = \gamma_A x - x^2\,, \quad x(0) = x \quad &\rightarrow \quad x(\tau_R) = \frac{x e^{-\gamma_A \tau_R}}{1 + \frac{x}{\gamma_A}\left(1 - e^{-\gamma_A \tau_R}\right)}\,,\,, \\
\partial_{\tau_R} y = \gamma_B y - y^2\,, \quad y(0) = y \quad &\rightarrow \quad y(\tau_R) = \frac{y e^{-\gamma_B \tau_R}}{1 + \frac{y}{\gamma_B}\left(1 - e^{-\gamma_B \tau_R}\right)}\,.
\end{aligned} \tag{B5}$$

Then $\Phi_{1,0}(\tau_R)$ satisfies a related differential equation,

$$\partial_{\tau_R}\Phi_{1,0} = x(\tau_R) + y(\tau_R)\,, \quad \Phi_{1,0}(0) = H_{1,0}(x,y,\tau)\,, \quad \Phi_{1,0}(\tau) = 0\,, \tag{B6}$$

whose solution is given by

$$\Phi_1(\tau_R) - \Phi_1(0) = \int_0^{\tau_R} d\tau_R' \left[x(\tau_R') + y(\tau_R')\right] \tag{B7}$$

The definition of $\Phi_{1,0}(\tau_R)$ in Eq. (B4) then implies that

$$H_{1,0}(x,y,\tau) = -\int_0^\tau d\tau_R[x(\tau_R)+y(\tau_R)] = \underbrace{-\log\left[1+\frac{x}{\gamma_A}\left(1-e^{-\gamma_A\tau}\right)\right]}_{H_A(x,\tau)} + \underbrace{-\log\left[1+\frac{y}{\gamma_B}\left(1-e^{-\gamma_B\tau}\right)\right]}_{H_B(y,\tau)} \tag{B8}$$

which yields Eq. (53) in the main text.

This same approach can be extended to higher orders of the series expansion Eq. (52). At the next order (θ^2), we have

$$\mathcal{L}H_{2,0} = -(x+y)H_{1,0}\,, \tag{B9}$$

which can again be solved by the method of characteristics. We define an analogous function,

$$\Phi_{2,0}(\tau_R) = H_{2,0}(x(\tau_R), y(\tau_R), \tau-\tau_R)\,, \tag{B10}$$

where the characteristic curves $x(\tau_R)$ and $y(\tau_R)$ are the same as above. Then $\Phi_{2,0}(\tau_R)$ satisfies

$$\partial_{\tau_R}\Phi_{2,0} = (x(\tau_R)+y(\tau_R))\Phi_{1,0}(\tau_R)\,, \quad \Phi_2(0) = H_{2,0}(x,y,\tau)\,, \quad \Phi_2(\tau) = 0\,, \tag{B11}$$

and hence

$$H_{2,0}(x,y,\tau) = -\int_0^\tau d\tau_R\,\Phi_{1,0}(\tau_R)\partial_{\tau_R}\Phi_{1,0}(\tau_R) = \frac{1}{2}\left[H_A(x,\tau)+H_B(y,\tau)\right]^2\,. \tag{B12}$$

Note that this $\mathcal{O}(\theta^2)$ term is independent of z, which means that it cannot contribute to averages involving f_{AB}.

The lowest order term in f_0 is $\mathcal{O}(\theta^2 f_0)$, since the $\mathcal{O}(\theta f_0)$ term vanishes. At this order, we have

$$\begin{aligned}\mathcal{L}H_{2,1} &= z\left[\frac{\partial H_{1,0}}{\partial x}+\frac{\partial H_{1,0}}{\partial y}-\rho\frac{\partial H_{2,0}}{\partial x\partial y}\right]+\rho(x+y)\frac{\partial H_{2,0}}{\partial x\partial y}\\ &= z\left[\frac{\partial H_A}{\partial x}+\frac{\partial H_B}{\partial y}-\rho\frac{\partial H_A}{\partial x}\frac{\partial H_B}{\partial y}\right]+\rho(x+y)\frac{\partial H_A}{\partial x}\frac{\partial H_B}{\partial y}\end{aligned} \tag{B13}$$

where H_A and H_B are defined as above. The solution to this equation proceeds in a similar fashion as above. Let $\Phi_{2,1}(\tau_R)$ be defined by

$$\Phi_{2,1}(\tau_R) = H_{2,1}(x(\tau_R), y(\tau_R), z(\tau_R), \tau_f-\tau_R)\,, \tag{B14}$$

where the characteristic curve z is defined by

$$\frac{\partial z(\tau_R)}{\partial\tau_R} = \gamma_{AB}z(\tau_R) - z(\tau_R)^2 + \rho(x(\tau_R)+y(\tau_R))\,, \quad z(0) = z\,. \tag{B15}$$

Then $\Phi_{2,1}(\tau_R)$ satisfies

$$\begin{aligned}&\frac{\partial\Phi_{2,1}}{\partial\tau_R} = -z(\tau_R)\left[\Phi_x(\tau_R)+\Phi_y(\tau_R)-\rho\Phi_x(\tau_R)\Phi_y(\tau_R)\right]-\rho\left[x(\tau_R)+y(\tau_R)\right]\Phi_x(\tau_R)\Phi_y(\tau_R)\,,\\ &\Phi_{2,1}(0) = H_{2,1}(x,y,z,\tau)\\ &\Phi_{2,1}(\tau) = 0\,,\end{aligned} \tag{B16}$$

where $\Phi_x(\tau_R)$ and $\Phi_y(\tau_R)$ are defined by

$$\begin{aligned}\Phi_x(\tau_R) &\equiv \left.\frac{\partial H_A}{\partial x}\right|_{x(\tau_R),\tau_f-\tau_R} &= -\frac{\left(1-e^{-\gamma_A\tau_f+\gamma_A\tau_R}\right)\left[\gamma_A+x\left(1-e^{-\gamma_A\tau_R}\right)\right]}{\gamma_A\left[\gamma_A+x\left(1-e^{-\gamma_A\tau_f}\right)\right]}\\ \Phi_y(\tau_R) &\equiv \left.\frac{\partial H_B}{\partial y}\right|_{y(\tau_R),\tau_f-\tau_R} &= -\frac{\left(1-e^{-\gamma_B\tau_f+\gamma_B\tau_R}\right)\left[\gamma_B+y\left(1-e^{-\gamma_B\tau_R}\right)\right]}{\gamma_B\left[\gamma_B+y\left(1-e^{-\gamma_B\tau_f}\right)\right]}\end{aligned} \tag{B17}$$

The solution for $H_{2,1}$ then follows as

$$\begin{aligned} H_{2,1}(x,y,z,\tau) = &\int_0^\tau d\tau_R \, z(\tau_R) \left[\Phi_x(\tau_R) + \Phi_y(\tau_R) - \rho\Phi_x(\tau_R)\Phi_y(\tau_R)\right] \\ &+ \underbrace{\int_0^\tau d\tau_R \, \rho[x(\tau_R) + y(\tau_R)]\Phi_x(\tau_R)\Phi_y(\tau_R)}_{\Upsilon(x,y,\tau)} \,, \end{aligned} \tag{B18}$$

where $\Upsilon(x,y,\tau)$ is a function that is independent of z. Combining this expression with the $H_{2,0}$ and $H_{1,0}$ terms above yields the perturbative solution in Eq. (56) in the main text. Thus, if we can find a solution for the characteristic curve $z(\tau_R)$ in Eq. (B15), then the leading order contributions to the generating function can be obtained by direct integration. We consider several such solutions in the Appendices below. To minimize confusion, we will use the notation $\psi(x,y,z,\tau_R)$ in place of $z(\tau_R)$ throughout the rest of this work, in order to emphasize the implicit dependence on the initial conditions $x(0) = x$, $y(0) = y$, and

Appendix C: Solution for nonrecombining loci

In the case of non-recombining loci ($\rho = 0$), the characteristic curve in Eq. (56b) reduces to a simple logistic equation, whose solution is given by

$$\psi(z,\tau') = \frac{z e^{-\gamma_{AB}\tau'}}{1 + \frac{z}{\gamma_{AB}}\left(1 - e^{-\gamma_{AB}\tau'}\right)} \,. \tag{C1}$$

This solution will also be valid for finite recombination rates, provided that $\rho \ll \gamma_A, \gamma_B, \gamma_\epsilon$. Substituting this solution into Eq. (59), we find that the equilibrium generating function can be expressed as a definite integral,

$$\begin{aligned} H(x,y,z) \approx\; & 1 - \theta\log\left(1 + \frac{x}{\gamma_A}\right) - \theta\log\left(1 + \frac{y}{\gamma_B}\right) + \frac{\theta^2}{2}\log^2\left(1 + \frac{x}{\gamma_A}\right) + \frac{\theta^2}{2}\log^2\left(1 + \frac{y}{\gamma_B}\right) \\ & + \theta^2\log\left(1 + \frac{x}{\gamma_A}\right)\log\left(1 + \frac{y}{\gamma_B}\right) - \theta^2 f_0 \int_0^\infty \frac{z e^{-\gamma_{AB}\tau'}}{1 + \frac{z}{\gamma_{AB}}(1 - e^{-\gamma_{AB}\tau'})} \\ & \times \left[\frac{\gamma_A + x(1 - e^{-\gamma_A\tau'})}{\gamma_A(\gamma_A + x)} + \frac{\gamma_B + y(1 - e^{-\gamma_B\tau'})}{\gamma_B(\gamma_B + y)}\right] d\tau' \,. \end{aligned} \tag{C2}$$

In this case, the integral can be evaluated using special functions,

$$\begin{aligned} & \int_0^\infty \frac{z e^{-\gamma_{AB}\tau'}}{1 + \frac{z}{\gamma_{AB}}(1 - e^{-\gamma_{AB}\tau'})} \left[\frac{\gamma_A + x(1 - e^{-\gamma_A\tau'})}{\gamma_A(\gamma_A + x)} + \frac{\gamma_B + y(1 - e^{-\gamma_B\tau'})}{\gamma_B(\gamma_B + y)}\right] d\tau' \\ & = \left(\frac{1}{\gamma_A} + \frac{1}{\gamma_B}\right)\log\left(1 + \frac{z}{\gamma_{AB}}\right) - \frac{xz\,{}_2F_1\left(1, 1, 2 + \frac{\gamma_A}{\gamma_{AB}}, -\frac{z}{\gamma_{AB}}\right)}{\gamma_A(x + \gamma_A)(\gamma_{AB} + \gamma_A)} - \frac{yz\,{}_2F_1\left(1, 1, 2 + \frac{\gamma_B}{\gamma_{AB}}, -\frac{z}{\gamma_{AB}}\right)}{\gamma_B(y + \gamma_B)(\gamma_{AB} + \gamma_B)} \,. \end{aligned} \tag{C3}$$

where ${}_2F_1(a,b,c,u)$ is the hypergeometric function. This integral simplifies even further in the limit that $\gamma_A = \gamma_B = 0$, where we find that

$$\begin{aligned} & \int_0^\infty \frac{z e^{-\gamma_{AB}\tau'}}{1 + \frac{z}{\gamma_{AB}}(1 - e^{-\gamma_{AB}\tau'})} \left[\frac{\gamma_A + x(1 - e^{-\gamma_A\tau'})}{\gamma_A(\gamma_A + x)} + \frac{\gamma_B + y(1 - e^{-\gamma_B\tau'})}{\gamma_B(\gamma_B + y)}\right] d\tau' \\ & \approx \frac{2}{\gamma_\epsilon}\mathrm{Li}_2\left(\frac{z}{\gamma_\epsilon + z}\right) + \left(\frac{1}{x} + \frac{1}{y}\right)\log\left(1 + \frac{z}{\gamma_\epsilon}\right) \,. \end{aligned} \tag{C4}$$

where $\mathrm{Li}_n(u)$ is the polylogarithm function. For $\gamma_\epsilon \ll 1$, the resulting distribution of f_{AB} is approximately uniformly distributed up to a cutoff around $\sim f_0$, with $f_{AB} \approx f_{aB} \approx 0$. This is much broader than the standard single-locus prediction, and reflects the dominant contribution of ancient nested mutations that originated long before the time of sampling.

We can also use this same solution to evaluate various moments $\Sigma_d^k(\gamma_A, \gamma_B, \gamma_\epsilon)$ directly. To do so, it will be helpful to make use of the following identities:

$$\frac{\partial \psi(z,\tau')}{\partial z} = \frac{e^{-\gamma_{AB}\tau'}}{\left[1 + \frac{z}{\gamma_{AB}}\left(1 - e^{-\gamma_{AB}\tau'}\right)\right]^2} \tag{C5}$$

and

$$\frac{\partial^k \psi(z,\tau')}{\partial z^k} = \frac{(-1)^{k+1}k!}{\gamma_{AB}^{k-1}} \frac{e^{-\gamma_{AB}\tau'}\left(1 - e^{-\gamma_{AB}\tau'}\right)^{k-1}}{\left[1 + \frac{z}{\gamma_{AB}}\left(1 - e^{-\gamma_{AB}\tau'}\right)\right]^{k+1}} \tag{C6}$$

for $k \geq 2$, such that

$$\psi_k(\tau') = \begin{cases} \frac{e^{-\gamma_{AB}\tau'}}{\left[1+\frac{2}{\gamma_{AB}}\left(1-e^{-\gamma_{AB}\tau'}\right)\right]^2} & \text{if } k = 1 \\ \frac{k!}{\gamma_{AB}^{k-1}} \frac{e^{-\gamma_{AB}\tau'}\left(1-e^{-\gamma_{AB}\tau'}\right)^{k-1}}{\left[1+\frac{2}{\gamma_{AB}}\left(1-e^{-\gamma_{AB}\tau'}\right)\right]^{k+1}} & \text{if } k \geq 2 \end{cases} \tag{C7}$$

Substituting these expressions into Eq. (60), we obtain

$$\Sigma_d^1 = -1 + \int_0^\infty \frac{e^{-\gamma_{AB}\tau'}}{\left[1 + \frac{2}{\gamma_{AB}}\left(1 - e^{-\gamma_{AB}\tau'}\right)\right]^2} \left[\frac{(1+\gamma_B)(\gamma_A + 1 - e^{-\gamma_A\tau'})}{\gamma_A} + \frac{(1+\gamma_A)(\gamma_B + 1 - e^{-\gamma_B\tau'})}{\gamma_B}\right] d\tau' \tag{C8}$$

$$\begin{aligned} &= \frac{\gamma_{AB} - \gamma_A - \gamma_B}{2 + \gamma_{AB}} + \frac{1+\gamma_B}{(2+\gamma_{AB})(\gamma_{AB}+\gamma_A)}\,{}_2F_1\left(1, 1, 2 + \frac{\gamma_A}{\gamma_{AB}}, -\frac{2}{\gamma_{AB}}\right) \\ &\quad + \frac{1+\gamma_A}{(2+\gamma_{AB})(\gamma_{AB}+\gamma_B)}\,{}_2F_1\left(1, 1, 2 + \frac{\gamma_B}{\gamma_{AB}}, -\frac{2}{\gamma_{AB}}\right) \end{aligned} \tag{C9}$$

and

$$\Sigma_d^2 = \int_0^\infty \frac{2}{\gamma_{AB}} \frac{e^{-\gamma_{AB}\tau'}\left(1 - e^{-\gamma_{AB}\tau'}\right)}{\left[1 + \frac{2}{\gamma_{AB}}\left(1 - e^{-\gamma_{AB}\tau'}\right)\right]^3} \left[\frac{(1+\gamma_B)(\gamma_A + 1 - e^{-\gamma_A\tau'})}{\gamma_A} + \frac{(1+\gamma_A)(\gamma_B + 1 - e^{-\gamma_B\tau'})}{\gamma_B}\right] d\tau' \tag{C10}$$

$$\begin{aligned} &= \frac{(1+\gamma_A)(1+\gamma_B)}{(2+\gamma_{AB})^3}\left[\frac{\gamma_A + 3\gamma_{AB} + \gamma_{AB}^2 + \gamma_{AB}\gamma_A}{(\gamma_A+\gamma_{AB})(1+\gamma_A)} + \frac{\gamma_B + 3\gamma_{AB} + \gamma_{AB}^2 + \gamma_{AB}\gamma_B}{(\gamma_B+\gamma_{AB})(1+\gamma_B)}\right] \\ &\quad + \frac{(1+\gamma_B)(4+\gamma_A+\gamma_{AB})\,{}_2F_1(1,1,3+\frac{\gamma_A}{\gamma_{AB}},\frac{-2}{\gamma_{AB}})}{(2+\gamma_{AB})^3(\gamma_A+2\gamma_{AB})} + \frac{(1+\gamma_A)(4+\gamma_B+\gamma_{AB})\,{}_2F_1(1,1,3+\frac{\gamma_B}{\gamma_{AB}},\frac{-2}{\gamma_{AB}})}{(2+\gamma_{AB})^3(\gamma_B+2\gamma_{AB})} \end{aligned} \tag{C11}$$

Appendix D: Solution for neutral loci

In the limit where γ_A, γ_B, γ_ϵ are all small compared to both 1 and ρ, we can obtain exact solutions for $F_k(\cdot)$ using special functions. Recall that since each of these scaled variables contains a factor of f_0, this neutral regime can apply even when the nominal fitness costs are much larger than $1/N$. When these conditions hold, the differential equation for the characteristic curve reduces to

$$\frac{\partial \psi(x,y,z,\tau')}{\partial \tau'} = -\rho\psi - \psi^2 + \frac{\rho x}{1 + x\tau'} + \frac{\rho y}{1 + y\tau'}\,, \quad \psi(x,y,z,0) = z\,. \tag{D1}$$

This equation is difficult to solve due to the presence of the time-dependent terms on the right-hand side, which vary over two different timescales $\sim 1/x$ and $\sim 1/y$. Note, however, that moments like $\Sigma_d^k(\rho)$ only depend on the value of this function in the special case that $x = y = 1$. Thus, for the purposes of computing $\Sigma_d^k(\rho)$, it will be sufficient to focus on the simpler equation

$$\frac{\partial \psi(z,\tau')}{\partial \tau'} = -\rho\psi - \psi^2 + \frac{2\rho}{1 + \tau'}\,, \quad \psi(z,0) = z\,, \tag{D2}$$

which can be non-dimensionalized using the transformation $u = (\tau+1)\rho$ and $\Psi = \psi/\rho$, yielding

$$\frac{\partial \Psi}{\partial u} = -\Psi - \Psi^2 + \frac{2}{u}\,, \quad \Psi(\rho) = \frac{z}{\rho} \tag{D3}$$

The general solution to this equation is of the form

$$\Psi(u) = \frac{\mathrm{Ei}(-u) + (1+u)^{-1} e^{-u} + c(z)}{\frac{u(2+u)}{2(1+u)}\mathrm{Ei}(-u) + \frac{1}{2}e^{-u} + \frac{u(2+u)}{2(1+u)}c(z)} \tag{D4}$$

or switching back to $\zeta = u - \rho = \rho\tau$,

$$\Psi(\zeta; z) = \frac{2(1+\rho+\zeta)\left[A(\zeta) + c(z)\right] + 2\rho e^{-\zeta}}{(\rho+\zeta)(2+\rho+\zeta)[A(\zeta)+c(z)] + (1+\rho+\zeta)\rho e^{-\zeta}} \tag{D5}$$

where we have defined

$$A(\zeta) \equiv \rho e^{\rho}\left[\mathrm{Ei}(-\rho-\zeta) - \mathrm{Ei}[-\rho)\right] \tag{D6}$$

and where the constant $c(z)$ is chosen to satisfy the initial condition $\Psi = z/\rho$ when $\zeta = 0$:

$$\frac{z}{\rho} = \frac{2(1+\rho)c(z) + 2\rho}{\rho(2+\rho)c(z) + (1+\rho)\rho} \tag{D7}$$

For our purposes, it will be useful to solve for $z = 2 + \epsilon$. Solving for $c(\epsilon)$, we obtain

$$c(\epsilon) = -\frac{1 + \left(\frac{1+\rho}{2}\right)\epsilon}{1 + \left(\frac{2+\rho}{2}\right)\epsilon} \tag{D8}$$

which has derivatives

$$\frac{\partial^k c(\epsilon)}{\partial \epsilon^k} = (-1)^{k-1} k! \left(\frac{2+\rho}{2}\right)^{k-1} \frac{1}{2} \frac{1}{\left(1 + \frac{2+\rho}{2}\epsilon\right)^{k+1}} \rightarrow (-1)^{k-1} k!\, (2+\rho)^{k-1}\, 2^{-k} \tag{D9}$$

It is then straightforward to compute derivatives of $\Psi(\zeta;\epsilon)$ with respect to ϵ. For the first derivative, we find,

$$\begin{aligned}\frac{\partial \Psi(\zeta;\epsilon)}{\partial \epsilon} &= \frac{2(1+\rho+\zeta)\frac{\partial c}{\partial \epsilon}}{(\rho+\zeta)(2+\rho+\zeta)[A(\zeta)+c(z)] + (1+\rho+\zeta)\rho e^{-\zeta}} \\ &\quad - \frac{\left[2(1+\rho+\zeta)[A(\zeta)+c(\epsilon)] + 2\rho e^{-\zeta}\right](\rho+\zeta)(2+\rho+\zeta)\frac{\partial c}{\partial \epsilon}}{\left[(\rho+\zeta)(2+\rho+\zeta)[A(\zeta)+c(z)] + (1+\rho+\zeta)\rho e^{-\zeta}\right]^2} && \text{(D10)} \\ &= \frac{2\rho e^{-\zeta}}{\left[(\rho+\zeta)(2+\rho+\zeta)[A(\zeta)+c(z)] + (1+\rho+\zeta)\rho e^{-\zeta}\right]^2}\left(\frac{\partial c}{\partial \epsilon}\right) && \text{(D11)}\end{aligned}$$

Higher derivatives are facilitated by writing this as

$$\frac{\partial \Psi(\zeta;\epsilon)}{\partial \epsilon} = \frac{N(\zeta)}{\left[D_1(\zeta)c(\epsilon) + D_2(\zeta)\right]^2}\left(\frac{\partial c}{\partial \epsilon}\right) \tag{D12}$$

where we have defined

$$\begin{aligned} N(\zeta) &= 2\rho e^{-\zeta} && \text{(D13)} \\ D_1(\zeta) &= (\rho+\zeta)(2+\rho+\zeta) && \text{(D14)} \\ D_2(\zeta) &= (\rho+\zeta)(2+\rho+\zeta)\rho e^{-\rho}\left[\mathrm{Ei}(-\rho-\zeta) - \mathrm{Ei}(-\rho)\right] + (1+\rho+\zeta)\rho e^{-\zeta} && \text{(D15)}\end{aligned}$$

The second, third, and fourth derivatives are therefore given by

$$\frac{\partial^2 \Psi}{\partial \epsilon^2} = \frac{N}{(D_1 c + D_2)^2}\left[\frac{\partial^2 c}{\partial \epsilon^2} - \frac{2D_1}{(D_1 c + D_2)}\left(\frac{\partial c}{\partial \epsilon}\right)^2\right] \tag{D16}$$

$$\frac{\partial^3 \Psi}{\partial \epsilon^3} = \frac{N}{(D_1 c + D_2)^2}\left[\frac{\partial^3 c}{\partial \epsilon^3} - \frac{6D_1}{(D_1 c + D_2)}\left(\frac{\partial^2 c}{\partial \epsilon^2}\right)\left(\frac{\partial c}{\partial \epsilon}\right) + \frac{6D_1^2}{(D_1 c + D_2)^2}\left(\frac{\partial c}{\partial \epsilon}\right)^3\right] \tag{D17}$$

$$\frac{\partial^4 \Psi}{\partial \epsilon^4} = \frac{N}{(D_1 c + D_2)^2}\left[\frac{\partial^4 c}{\partial \epsilon^4} - \frac{D_1}{(D_1 c + D_2)}\left[8\left(\frac{\partial^3 c}{\partial \epsilon^3}\right)\left(\frac{\partial c}{\partial \epsilon}\right) + 6\left(\frac{\partial^2 c}{\partial \epsilon^2}\right)^2\right]\right. \tag{D18}$$

$$\left. + \frac{36D_1^2}{(D_1 c + D_2)^2}\left(\frac{\partial^2 c}{\partial \epsilon^2}\right)\left(\frac{\partial c}{\partial \epsilon}\right)^2 - \frac{24D_1^3}{(D_1 c + D_2)^3}\left(\frac{\partial c}{\partial \epsilon}\right)^4\right] \tag{D19}$$

Evaluating at $\epsilon = 0$, we have

$$\left.\frac{\partial \Psi}{\partial \epsilon}\right|_{\epsilon=0} = \frac{N}{2(D_1 c(0) + D_2)^2} \tag{D20}$$

$$\left.\frac{\partial^2 \Psi}{\partial \epsilon^2}\right|_{\epsilon=0} = \frac{-N}{2(D_1 c(0) + D_2)^2}\left[(2+\rho) + \frac{D_1}{D_1 c_0 + D_2}\right] \tag{D21}$$

$$\left.\frac{\partial^4 \Psi}{\partial \epsilon^4}\right|_{\epsilon=0} = \frac{-3N}{2(D_1 c(0) + D_2)^2}\left[(2+\rho)^3 + \frac{3D_1(2+\rho)^2}{(D_1 c_0 + D_2)} + \frac{3D_1^2(2+\rho)}{(D_1 c_0 + D_2)^2} + \frac{D_1^3}{(D_1 c_0 + D_2)^3}\right] \tag{D22}$$

This yields

$$\Sigma_d^1(\rho) = -1 + \int_0^\infty \frac{e^{-\zeta}(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)^2}\, d\zeta \tag{D23}$$

$$\Sigma_d^2(\rho) = \int_0^\infty \frac{e^{-\zeta}(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)^2}\left[(2+\rho) + \frac{(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)}\right] d\zeta \tag{D24}$$

$$\begin{aligned}\Sigma_d^4(\rho) = \int_0^\infty \frac{e^{-\zeta}(\rho+\zeta)(2+\rho+\zeta)}{D(\zeta)^2}&\left[(2+\rho)^3 + \frac{3(\rho+\zeta)(2+\rho+\zeta)(2+\rho)^2}{D(\zeta)}\right.\\ &\left. + \frac{3[(\rho+\zeta)(2+\rho+\zeta)]^2(2+\rho)}{D(\zeta)^2} + \frac{3[(\rho+\zeta)(2+\rho+\zeta)]^3}{D(\zeta)^3}\right] d\zeta\end{aligned} \tag{D25}$$

where we have defined

$$D(\zeta) = (\rho+\zeta)(2+\rho+\zeta)[-1 + \rho e^{\rho}[\mathrm{Ei}(-\rho-\zeta) - \mathrm{Ei}(-\rho)] + (1+\rho+\zeta)\rho e^{-\zeta} \tag{D26}$$

Appendix E: Solution for strong selection or recombination

We finally consider the regime where $\gamma_{AB} \gg 1$, which can occur either when any of γ_A, γ_B, γ_ϵ or ρ are large compared to one. In this limit, we can obtain solutions via perturbation theory, treating the ψ^2 term as a small correction. We first rescale time $u = \tau\gamma_{AB}$, so that

$$\frac{\partial \psi}{\partial u} = -\psi - \frac{1}{\gamma_{AB}}\psi^2 + \frac{\rho}{\gamma_{AB}}\left[\frac{e^{-\gamma_A u/\gamma_{AB}}}{1 + \frac{1}{\gamma_A}\left(1 - e^{-\gamma_A u/\gamma_{AB}}\right)} + \frac{e^{-\gamma_B u/\gamma_{AB}}}{1 + \frac{1}{\gamma_B}\left(1 - e^{-\gamma_B u/\gamma_{AB}}\right)}\right] \tag{E1}$$

We can put this in a more compact form by defining $\epsilon = 1/\gamma_{AB}$, $\alpha = \rho/\gamma_{AB}$, and $f(u) = x(u) + y(u)$, so that

$$\frac{\partial \psi}{\partial u} = -\psi - \epsilon\psi^2 + \alpha f(u) \tag{E2}$$

with initial condition $\psi(0) = z$. We can solve this equation using a perturbation expansion in ϵ, defining

$$\psi(u) = \sum_{k=0}^{\infty} \epsilon^k \psi_k(u) \tag{E3}$$

$$f(u) = \sum_{k=0}^{\infty} \epsilon^k f_k(u) \tag{E4}$$

with $\psi_k(0) = z\delta_{k,0}$. At zeroth order, we have

$$\frac{\partial \psi_0}{\partial u} = -\psi_0 + \alpha f_0(u) \tag{E5}$$

and hence

$$\psi_0(u;z) = ze^{-u} + \alpha e^{-u} \int_0^u e^{u'} f_0(u')\, du' \tag{E6}$$

For our purposes, it will suffice to continue this formal solution through second order in ϵ. At first order, we have

$$\frac{\partial \psi_1}{\partial u} = -\psi_1 + \alpha f_1(u) - \psi_0(u;z)^2 \tag{E7}$$

and hence

$$\psi_1(u;z) = \alpha e^{-u} \int_0^u e^{u'} f_1(u')\, du' - e^{-u} \int_0^u e^{u'} \psi_0(u';z)^2\, du' \tag{E8}$$

$$= \alpha e^{-u} \int_0^u e^{u'} f_1(u')\, du' - z^2 e^{-u}(1 - e^{-u}) + \dots \tag{E9}$$

At second order, we have

$$\frac{\partial \psi_2}{\partial u} = -\psi_2 + \alpha f_2(u) - 2\psi_0(u;z)\psi_1(u,;z) \tag{E10}$$

and hence

$$\psi_2(u;z) = \alpha e^{-u} \int_0^u e^{u'} f_2(u')\, du' - e^{-u} \int_0^u e^{u'} 2\psi_0(u';z)\psi_1(u;z)\, du' \tag{E11}$$

$$= z^2 e^{-u} \left(1 - e^{-u}\right)^2 + \dots \tag{E12}$$

Finally, at third order, we have

$$\frac{\partial \psi_3}{\partial u} = -\psi_3 + \alpha f_3(u) - \psi_1(u;z)^2 - 2\psi_0(u;z)\psi_2(u;z) \tag{E13}$$

and hence

$$\psi_3(u;z) = \alpha e^{-u} \int_0^u e^{u'} f_3(u')\, du' - e^{-u} \int_0^u e^{u'} [\psi_1(u';z)^2 + 2\psi_0(u';z)\psi_2(u';z)\, du' \tag{E14}$$

$$= -z^4 e^{-u}(1 - e^{-u})^3 + \dots \tag{E15}$$

We can use these formal solutions to compute derivatives with respect to z:

$$\left.\frac{\partial \psi}{\partial z}\right|_{z=2} = \left.\frac{\partial \psi_0}{\partial z}\right|_{z=2} + \mathcal{O}(\epsilon) \approx e^{-u} \tag{E16}$$

$$\left.\frac{\partial^2 \psi}{\partial z^2}\right|_{z=2} = \epsilon \left.\frac{\partial^2 \psi_1}{\partial z^2}\right|_{z=2} + \mathcal{O}(\epsilon^2) \approx -\epsilon e^{-u} \int_0^u e^{u'} \left[2e^{-2u'}\right] = -2\epsilon e^{-u}(1 - e^{-u}) \tag{E17}$$

$$\left.\frac{\partial^4 \psi}{\partial z^4}\right|_{z=2} = \epsilon^3 \left.\frac{\partial^4 \psi_3}{\partial z^4}\right|_{z=2} + \mathcal{O}(\epsilon^4) \approx= -24\epsilon^3 e^{-u}(1 - e^{-u})^3 \tag{E18}$$

Note that all three expressions are independent of $f(u)$ at lowest order. We can then use these expressions to calculate LD statistics:

$$\begin{aligned}\Sigma_d^1(\gamma_A,\gamma_B,\gamma_{AB},\rho) = -1 + \frac{1}{\gamma_{AB}}\int_0^\infty du\, e^{-u}\Bigg[&(\gamma_B+1)\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\\ &+(\gamma_A+1)\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}+\rho\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}\Bigg]\end{aligned} \tag{E19}$$

and

$$\begin{aligned}\Sigma_d^2(\gamma_A,\gamma_B,\gamma_{AB},\rho) = \frac{1}{\gamma_{AB}^2}\int_0^\infty du\, 2e^{-u}(1-e^{-u})\Bigg[&(\gamma_B+1)\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\\ &+(\gamma_A+1)\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}+\rho\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}\Bigg]\end{aligned} \tag{E20}$$

and

$$\begin{aligned}\Sigma_d^4(\gamma_A,\gamma_B,\gamma_{AB},\rho) = \frac{(\gamma_A+1)(\gamma_B+1)}{\gamma_{AB}^4}\int_0^\infty du\, 4e^{-u}(1-e^{-u})^3\Bigg[&(\gamma_B+1)\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\\ &+(\gamma_A+1)\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}+\rho\cdot\frac{\gamma_A+1-e^{-\gamma_A u/\gamma_{AB}}}{\gamma_A}\cdot\frac{\gamma_B+1-e^{-\gamma_B u/\gamma_{AB}}}{\gamma_B}\Bigg]\end{aligned} \tag{E21}$$

In the limit that $\gamma_A, \gamma_B \gg 1$, this reduces to Eq. (66) in the main text.

Appendix F: Transition to the Quasi-Linkage Equilibrium (QLE) regime

Using the quasi-stationary distribution in Eq. (68) in the main text, it is straightforward to show that the first several conditional averages are given by

$$\langle D|f_{Ab},f_{aB}\rangle = \langle f_{AB}|f_{Ab},f_{aB}\rangle - f_A f_B \approx 0 \tag{F1}$$

$$\langle D^2|f_{Ab},f_{aB}\rangle = \langle (f_{AB}-f_Af_B)^2|f_{Ab},f_{aB}\rangle \approx \frac{2NRf_{Ab}f_{aB}}{(2NR)^2} \tag{F2}$$

and

$$\begin{aligned}\langle D^4|f_{Ab},f_{aB}\rangle &\approx \langle f_{AB}^4|f_{Ab},f_{aB}\rangle - 4\langle f_{AB}^3|f_Af_B\rangle f_Af_B + 6\langle f_{AB}^2|f_{Ab},f_{aB}\rangle (f_Af_B)^2 - 4\langle f_{AB}\rangle (f_Af_B)^3 + (f_Af_B)^4\\ &= \frac{(2NRf_Af_B+3)(2NRf_Af_B+2)(2NRf_Af_B+1)f_Af_B}{(2NR)^3} \qquad \text{(F3)}\\ & \qquad \text{(F4)}\end{aligned}$$

in the limit that $\rho \gg 1$ and $f_A, f_B \ll 1$. Averaging over the slowly evolving f_{Ab} and f_{aB} frequencies then yields the moments in Eq. (69) in the main text. The ratio between the fourth and second moments then follows as

$$\eta(f_0) = \frac{\sigma_d^4(f_0)}{3\sigma_d^2(f_0)} = \frac{1+NRf_0^2}{2(NR)^2} \tag{F5}$$

To obtain a formula for η that works throughout the full range of NR and f_0 values, we asymptotically match this expression with the corresponding formula from Eq. (60), which yields

$$\eta(NR,f_0) = \left(\frac{1+NRf_0^2}{2(NR)^2}\right)e^{-\frac{1}{2NRf_0^2}} + \left(\frac{f_0\Sigma_d^4(2NRf_0)}{3\Sigma_d^2(2NRf_0)}\right)\left(1-e^{-\frac{1}{2NRf_0^2}}\right) \tag{F6}$$

where $\Sigma_d^2(\rho)$ and $\Sigma_d^4(\rho)$ are defined by Eqs. (D24) and (D25) above.

Appendix G: Estimating frequency-resolved LD in finite samples

As described in the main text, we can obtain unbiased finite-sample estimators for $\sigma_d^k(f_0)$ by expanding the D and $f_A(1-f_A)f_B(1-f_B)$ terms in Eq. (7) and applying the moment formula in Eq. (77). To ensure a smooth mapping to the $f_0 \to \infty$ limit, it is useful to define a modified $M(\vec{n})$ function,

$$M_{i,j,k,l}(\vec{n}) = \left[\frac{n_{Ab}! \left(1-\frac{1}{nf_0}\right)^{n_{Ab}-i}}{n^i (n_{Ab}-i)!} \cdot \frac{n_{aB}! \left(1-\frac{1}{nf_0}\right)^{n_{aB}-i}}{n^j (n_{aB}-j)!} \cdot \frac{n_{AB}! \left(1-\frac{2}{nf_0}\right)^{n_{AB}-k}}{n^k (n_{AB}-k)!} \cdot \frac{n_{ab}!}{n^l (n_{ab}-l)!} \right] , \tag{G1}$$

which satisfies a related moment formula

$$\langle M_{i,j,k,l}(\vec{n}) \rangle \approx \left\langle f_{Ab}^i f_{aB}^j f_{AB}^k f_{ab}^l e^{-f_A/f_0 - f_B/f_0} \right\rangle , \tag{G2}$$

in the limits that $f_0 \ll 1$ or $f_0 \gg 1$. Carrying out this procedure for the first few moments, we obtain

$$\left\langle D e^{-\frac{f_A+f_B}{f_0}} \right\rangle = \langle M_{0,0,1,1}(\vec{n}) - M_{1,1,0,0}(\vec{n}) \rangle , \tag{G3a}$$

$$\left\langle D^2 e^{-\frac{f_A+f_B}{f_0}} \right\rangle = \langle M_{0,0,2,2}(\vec{n}) - 2M_{1,1,1,1}(\vec{n}) + M_{2,2,0,0}(\vec{n}) \rangle , \tag{G3b}$$

and

$$\begin{aligned} \left\langle f_A(1-f_A) f_B(1-f_B) e^{-\frac{f_A+f_B}{f_0}} \right\rangle = \langle & M_{2,2,0,0}(\vec{n}) + M_{1,2,0,1}(\vec{n}) + M_{2,1,1,0}(\vec{n}) + M_{1,1,1,1}(\vec{n}) \\ & + M_{2,1,0,1}(\vec{n}) + M_{1,1,0,2}(\vec{n}) + M_{2,0,1,1}(\vec{n}) + M_{1,0,1,2}(\vec{n}) + M_{1,2,1,0}(\vec{n}) + M_{0,2,1,1}(\vec{n}) \\ & + M_{1,1,2,0}(\vec{n}) + M_{0,1,2,1}(\vec{n}) + M_{1,1,1,1}(\vec{n}) + M_{0,1,1,2}(\vec{n}) + M_{1,0,1,2}(\vec{n}) + M_{0,0,2,2}(\vec{n}) \rangle , \end{aligned} \tag{G3c}$$

which can be combined to obtain count-based formulae for $\sigma_d^1(f_0)$ and $\sigma_d^2(f_0)$. One can then apply these estimators to genomic data by replacing the ensemble averages with appropriate sums over many functionally similar pairs of sites.

Appendix H: Applications to polymorphism data from E. rectale

The linkage disequilibrium estimates in Fig. 7 were obtained from a sample of $n = 109$ *E. rectale* genomes that my collaborators and I analyzed in a previous study (Garud et al., 2019). In that work, we used a referenced-based approach to identify single nucleotide variants (SNVs) in the intra-host populations of several common species of gut bacteria from a panel of ∼1000 sequenced fecal samples. We also identified samples in which the haplotype of the dominant strain of a given could be resolved with high confidence. This analysis yielded 159 "quasi-phaseable" *E. rectale* genomes, which we used to identify between-host SNVs in the global *E. rectale* population. After controlling for population structure, we identified a subset of $n = 109$ samples that were inferred to descend from the largest clade. These 109 genomes were used as the basis for the analysis in Fig. 7 in the present work. These calculations started from the same collection of SNVs identified in Garud et al. (2019), and focused specifically on the subset of SNVs that were located at fourfold degenerate sites in core genes. Two-site haplotypes and coordinate distances were recorded for all pairs of SNVs located within 4 consecutive genes of each other on the *E. rectale* reference genome, and the A and B alleles were defined to coincide with the minority allele at each site. As a control, analogous two-site haplotypes were recorded for pairs of SNVs from a large number of randomly selected genes. The full collection of two-site haplotypes is provided in Supplementary Data. These counts were used as inputs for each of the calculations described in Fig. 7.